1-2014

Specification Testing for Transformation Models

Arthur LEWBEL
Boston College

Xun LU
Hong Kong University of Science and Technology

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Specification Testing for Transformation Models with an Application to Generalized Accelerated Failure-time Models*

Arthur Lewbel,[a] Xun Lu[b] and Liangjun Su[c]
[a] Department of Economics, Boston College
[b] Department of Economics, Hong Kong University of Science and Technology
[c] School of Economics, Singapore Management University

January 31, 2014

Abstract

Consider a nonseparable model $Y = R(X, U)$ where Y and X are observed, while U is unobserved and conditionally independent of X. This paper provides the first nonparametric test of whether R takes the form of a transformation model, meaning that Y is monotonic in the sum of a function of X plus a function of U. Transformation models of this form are commonly assumed in economics, including, e.g., standard specifications of duration models and hedonic pricing models. Our test statistic is asymptotically normal under local alternatives and consistent against nonparametric alternatives violating the implied restriction. Monte Carlo experiments show that our test performs well in finite samples. We apply our results to test for specifications of generalized accelerated failure-time (GAFT) models of the duration of strikes.

Keywords: additivity, control variable, endogenous variable, monotonicity, nonparametric nonseparable model, hazard model, specification test, transformation model, unobserved heterogeneity

JEL Classification: C12, C14

*Halbert White inspired this project, brought us together to work on it, and provided substantial advice, discussion, and enthusiasm. We deeply mourn his passing. We gratefully thank the co-editor Jianqinq Fan, the associate editor, and three anonymous referees for their many constructive comments. We are also indebted to Songnian Chen, Yanqin Fan, and Shakeeb Khan for helpful comments and suggestions. Lu acknowledges support from Hong Kong University of Science and Technology under grant number FSECS13BM02. Su acknowledges support from the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021.

1 Introduction

We consider a general nonseparable structural equation

$$Y = R(X, U), \tag{1.1}$$

where Y is a scalar observable outcome, X a $d_x \times 1$ vector of observable covariates of interest, U a $d_u \times 1$ vector of unobservable causes or errors, and R an unknown measurable function. Our goal is to test the following hypotheses:[1]

$\mathbb{H}_{10}$: There exist three measurable functions $G : \mathbb{R} \to \mathbb{R}$, $H_1 : \mathbb{R}^{d_x} \to \mathbb{R}$ and $H_2 : \mathbb{R}^{d_u} \to \mathbb{R}$ such that $Y = G[H_1(X) + H_2(U)]$ a.s., and G is strictly monotonic.

$\mathbb{H}_{1A}$: $\mathbb{H}_{10}$ is false.

Specifications that are monotonic functions of additive models have been called "transformation models" (e.g., Chiappori et al., 2013), or "transformed additively separable models" (e.g., Jacho-Chávez et al., 2010), or "generalized additive models with unknown link function" (e.g., Horowitz, 2001, and Horowitz and Mammen, 2004).

Broadly speaking, there are two kinds of transformation models that are common in the economics literature. The first type assumes that Y and X are observable, U is unobservable, and the link function $G(\cdot)$ may be known or unknown. Our paper belongs to this category. Ridder (1990), Horowitz (1996), Ekeland et al. (2004), Ichimura and Lee (2011), and Chiappori et al. (2013) discuss identification and estimation for transformation models of this category. Since U is unobservable in this class of models, only the functions G and H_1 are identified and estimated. In the second type of transformation model, both X and U are observable, and Y is an object that can be estimated such as a conditional mean or quantile function. Horowitz (2001), Horowitz and Mammen (2004, 2007, 2011), Horowitz and Lee (2005), and Jacho-Chávez et al. (2010) provide identification and estimation results for this second kind of transformation model, while Gozalo and Linton (2001) consider specification tests for such models. See also Horowitz (2014) for a recent survey on the latter class of models.

The transformation models under our null are commonly used (and hence assumed to hold) in a wide range of economic applications. For example, they are often used to study duration data (see, e.g., Heckman and Singer, 1984, Keifer, 1988, Mata and Portugal, 1994, Engle, 2000, and Abbring et al., 2008), including generalized accelerated failure-time (GAFT) models, which includes accelerated failure-time (AFT) models, proportional hazard (PH) models, and mixed proportional hazard (MPH) models as special cases. The MPH specification in particular is a widely used class of duration data specifications (for a review, see Van den Berg, 2001).

Despite its popularity, economic theory rarely justifies the MPH specification. For example, Van den Berg (2001, p. 3400) points out that "the MPH model specification is not derived from economic theory and it remains to be seen whether the MPH specification is actually able to capture important theoretical relations." He also provides many specific economic examples where the MPH specification is violated. In their microeconometrics textbook, Cameron and Trivedi (2005, p. 613) say that "the multiplicative

[1] Under the null $\mathbb{H}_{10}$, the function H_2 is generally not identified and one could simply replace $H_2(U)$ by a scalar unobservable $\tilde{U} \equiv H_2(U)$. We maintain the notation $H_2(U)$ to emphasize that, under the alternative, Y can depend on multi-dimensional unobservables as in some common nonseparable models (e.g., random coefficients models). As our notation shows, our test allows for the possibility that, under the alternative, there need not exist a function of U like $H_2(U)$ such that $R(X, U)$ depends on U only through $H_2(U)$.

heterogeneity assumption [in MPH models] is also rather special, but it is mathematically convenient..." Given the popularity (and the limitations) of GAFT models, especially MPH models, it is obvious that a formal specification test of these models would be useful for empirical research. While some specification tests for certain parametric forms of duration models exist (see, e.g., Fernandes and Grammig, 2005), to the best of our knowledge, ours is the first that specifically tests for the testable implications of the general specification of GAFT models.[2]

Another major set of applications of transformation model specifications where U is unobservable are hedonic models (see, e.g., Ekeland et al., 2004, and Heckman et al., 2005). Here again, we believe that our paper is the first to provide a general specification test for this class of transformation models.

A conditional exogeneity assumption is imposed to test $\mathbb{H}_{10}$, i.e., we assume that U and X are conditionally independent, conditioning on an observable covariate vector Z. This is analogous to the conditional unconfoundedness assumption in the treatment effect literature, and to the assumptions required for use of control function type methods of dealing with endogeneity (see, e.g., Heckman and Robb 1986, and Blundell and Powell, 2003). Chiappori et al. (2013) provide a nonparametric estimator for the transformation model under similar assumptions.

We first show that if the data are generated by a transformation model, i.e., $\mathbb{H}_{10}$ holds, then the ratio of the derivatives with respect to Y and to X of the conditional CDF of Y given (X, Z) can be written as a product of functions of X and Y.[3] We then use local polynomial methods to estimate these derivatives, and construct test statistics based on the L_2 distance between restricted and unrestricted estimators of this ratio of derivatives. We show that our test statistic is asymptotically normal under the null and under a sequence of Pitman local alternatives and is consistent against the alternatives violating the implied restriction. To facilitate the application of our test, we propose to use subsampling to obtain the p-values or critical values. We also evaluate our test both in a Monte Carlo setting, and in an empirical application concerning duration of strikes by manufacturing workers.

Our null $\mathbb{H}_{10}$ is weaker than additive separability but stronger than monotonicity. Lu and White (2013) and Su et al. (2013) propose tests for additive separability under the same conditional exogeneity assumption we make, i.e., they test whether there exist two unknown measurable functions G_1 and G_2 such that

$$Y = G_1(X) + G_2(U) \text{ a.s.}$$

Testing $\mathbb{H}_{10}$ is more general than testing for separability, since our null is equivalent to additive separability in the special case where G is known to be the identity function. Hence if we reject $\mathbb{H}_{10}$, then we also reject additive separability.

Hoderlein et al. (2011) (HSW) test for monotonicity under a conditional exogeneity assumption. Let $\tilde{U} \equiv H_2(U)$. HSW test whether there exists a function $\tilde{R}$ such that

$$Y = \tilde{R}(X, \tilde{U})$$

where $\tilde{R}$ is strictly monotonic in its second argument. Our null is stronger than monotonicity, so if the HSW test rejects monotonicity, then our null $\mathbb{H}_{10}$ is also rejected. Our null $\mathbb{H}_{10}$ combines monotonicity

[2] Recently, Chiappori et al. (2013) provide a nonparametric test, not for the transformation model specification itself, but for a conditional exogeneity assumption within the context of a transformation model. Still, their test might be interpreted as a model specification test. See Remark 2.6 in Section 2.1 for details.

[3] Horowitz (1996) considers the estimation of the semiparametric model under our null, where H_1 takes a parametric form (unlike our nonparametric case) and without covariates Z. His estimator also relies on the implication that the ratio of the derivatives is a multiplicative function of X and Y.

with the additional restriction that the observable X and unobservable $\tilde{U}$ are additively separable under a transformation function G. Our test exploits this additivity restriction, and so should be generally stronger than HSW for testing $\mathbb{H}_{10}$. Also, the HSW test requires that Z not be empty, while our test of $\mathbb{H}_{10}$ can be applied even if we have no conditioning covariate Z.

Note that in all these models, under the null Y equals a function of X and a scalar unobservable $\tilde{U}$, e.g., $\tilde{U} \equiv H_2(U)$ or $\tilde{U} \equiv G_2(U)$, but under the alternative U may be a random vector.

The rest of the paper is organized as follows. In Section 2, we propose and motivate our test. In Section 3, we show that our test statistics are asymptotically normal under the null, and we analyze their global and local power. In Section 4, we conduct some Monte Carlo simulations to evaluate the finite sample performance of our test statistics. In Section 5, we provide an empirical application to testing for the specification of GAFT models in data on the durations of strikes. In Section 6, we discuss extensions to other closely related hypotheses. Section 7 concludes, and mathematical proofs are relegated to the Appendix.

2 A Specification Test for Transformation Models

In this section, we describe implications of $\mathbb{H}_{10}$ that are used to motivate our test construction, and then describe our proposed test statistic.

2.1 Motivation

To construct our test, we first impose a conditional exogeneity assumption. Let $X \perp U \mid Z$ denote that X and U are independent given Z.

Assumption A.1. Let Z be an observable random vector of dimension $d_z \in \mathbb{N}$, such that $X \perp U \mid Z$ and that X and U are not measurable with respect to the sigma-field generated by Z.

Assumption A.1 is equivalent to the unconfoundedness assumption in the treatment effect literature and is widely used to identify causal effects. For detailed discussions, see Altonji and Matzkin (2005), Hoderlein and Mammen (2007), Imbens and Newey (2009), and White and Lu (2011), among others. It is also closely related to the assumptions used to allow for endogeneity in the control function literature, where Z would equal the residuals from a regression of X on exogenous instruments. See, e.g., Heckman and Robb (1986) and Blundell and Powell (2003, 2004). Under $\mathbb{H}_{10}$, the condition $X \perp U \mid Z$ can be relaxed a bit to $X \perp H_2(U) \mid Z$ in Theorem 2.1(a) below.

Let $F(\cdot \mid x, z) \equiv F_{Y|X,Z}(\cdot \mid x, z)$ and $f(\cdot \mid x, z) \equiv f_{Y|X,Z}(\cdot \mid x, z)$ denote the conditional cumulative distribution function (CDF) and probability density function (PDF) of Y given $(X, Z) = (x, z)$, respectively. Let $V = (X', Z')'$. Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{V}$, and $\mathcal{W}$ denote the supports of X, Y, V, and W, respectively. Note that we allow the support of $F(\cdot|x, z)$ to change according to the values x and z. Let $r^0(y; x, z) \equiv \frac{D_x F(y|x,z)}{f(y|x,z)}$, so $r^0(y; x, z)$ is the ratio of two partial derivatives of $F(y \mid x, z)$, since $f(y \mid x, z) = \partial F(y \mid x, z)/\partial y$ and $D_x F(y \mid x, z) \equiv \partial F(y \mid x, z)/\partial x$.

The following theorem characterizes some useful properties of the transformation model under $\mathbb{H}_{10}$.

Theorem 2.1 *Suppose that $f(y \mid x, z) \neq 0$ for all $(y, x, z) \in \mathcal{W}$.*

(a) If $\mathbb{H}_{10}$ and A.1 hold and the first order (partial) derivatives of G and H_1 exist, then there exist two measurable functions $s_1 : \mathbb{R}^{d_x} \to \mathbb{R}^{d_x}$ and $s_2 : \mathbb{R} \to \mathbb{R}_+$ (or $s_2 : \mathbb{R} \to \mathbb{R}_-$) such that

$$r^0(Y; X, Z) = s_1(X)\, s_2(Y) \ \ a.s., \tag{2.1}$$

where $s_1(x) = -\partial S_1(x)/\partial x$ for some measurable function $S_1 : \mathbb{R}^{d_x} \to \mathbb{R}$, and $1/s_2(y) = \partial S_2(y)/\partial y$ for some measurable function $S_2 : \mathbb{R} \to \mathbb{R}$.

(b) If there exist two measurable functions $s_1 : \mathbb{R}^{d_x} \to \mathbb{R}^{d_x}$ and $s_2 : \mathbb{R} \to \mathbb{R}_+$ (or $s_2 : \mathbb{R} \to \mathbb{R}_-$) such that (2.1) holds, $s_1(x) = -\partial S_1(x)/\partial x$ for some function $S_1 : \mathbb{R}^{d_x} \to \mathbb{R}$, and $1/s_2(y) = \partial S_2(y)/\partial y$ for some measurable function $S_2 : \mathbb{R} \to \mathbb{R}$, then $\mathbb{H}_{10}$ holds in the sense that there exist two measurable functions $G : \mathbb{R} \to \mathbb{R}$ and $H_1 : \mathbb{R}^{d_x} \to \mathbb{R}$ such that

$$Y = G\left[H_1(X) + \tilde{U}\right] \ \ a.s. \tag{2.2}$$

where G is strictly monotonic and differentiable, all first order partial derivatives of H_1 exist, and $\tilde{U}$ is a scalar unobservable random variable satisfying $X \perp \tilde{U} \mid Z$.

Remark 2.1. Theorem 2.1(a) says that under $\mathbb{H}_{10}$ and the conditional exogeneity condition in A.1, the ratio $r^0(y; x, z)$ is free of z and can be factored into the product of a function s_1 of x and a function s_2 of y, where the function s_1 can be written as the derivative of a scalar function, and the function s_2 does not alternate in sign on its support. Theorem 2.1(b) says the converse is also true: as long as the factorization in (2.1) holds with s_1 and s_2 satisfying appropriate conditions, the observables (Y, X, Z) will satisfy the version of transformation model (2.2) under the null. Note that even though U can be a vector in the true data generating process, given the conditions of Theorem 2.1(b), there exists a scalar unobservable $\tilde{U}$ that satisfies equation (2.2) and the conditional exogeneity in A.1.

Remark 2.2. Theorem 2.1 gives a characterization of $\mathbb{H}_{10}$, but it does not by itself provide a test for $\mathbb{H}_{10}$. The proof of Theorem 2.1(a) shows that s_1 and s_2 in the theorem depend on the unknown functions H_1 and G, respectively, so we cannot directly test equation (2.1). We instead propose a feasible and straightforward test statistic that is based on implications of the factorization in (2.1).

Let $\mathcal{Y}_0 \equiv [\underline{y}, \bar{y}] \subset \mathcal{Y}$ for finite real numbers $\underline{y}$ and $\bar{y}$. Let $\mathbf{1}\{\cdot\}$ denote the indicator function that equals one when $\cdot$ is true and zero otherwise, and let $E_Y(\cdot)$ and $E_{XZ}(\cdot)$ denote expectations with respect to Y and (X, Z), respectively. Define

$$\begin{aligned} r(y; x, z) &\equiv r^0(y; x, z)\, \mathbf{1}\{y \in \mathcal{Y}_0\}, \\ r_0 &\equiv E_Y E_{XZ}[r(Y; X, Z)], \\ r_1(x) &\equiv E[r(Y; x, Z)], \\ r_2(y) &\equiv E[r(y; X, Z)], \end{aligned} \tag{2.3}$$

where $r(y; x, z)$ denotes a trimmed version of $r^0(y; x, z)$. Note that r, r_0, r_1 and r_2 are all $d_x \times 1$ vectors. The following corollary summarizes a testable implication of (2.1) under $\mathbb{H}_{10}$ and A.1.

Corollary 2.2 *Suppose that $\mathbb{H}_{10}$ and A.1 hold. Then*

$$r(Y; X, Z) \circ r_0 = r_1(X) \circ r_2(Y) \ \ a.s., \tag{2.4}$$

where $\circ$ denotes the Hadamard product.

Remark 2.3. This corollary remains valid if we drop the indicator $\mathbf{1}\{y \in \mathcal{Y}_0\}$ in the definition of r in (2.3). Equivalently, one can take $\mathcal{Y}_0 = \mathcal{Y}$ in the definition of r and still obtain the above result provided that r is well defined. We incorporate the indicator function in our theorem to permit the trimming of the data in the tails that facilitates the rigorous establishment of the asymptotic properties of our test. Specifically our asymptotic theory below requires consistent estimation of $r(y; x, z)$ uniformly in $(y; x, z) \in \mathcal{W}_0 \equiv \mathcal{Y}_0 \times \mathcal{V}$. If $f(y \mid x, z)$ is too close to zero for some values of $(y, x, z) \in \mathcal{W}$, then we cannot estimate $r(y; x, z)$ uniformly in $(y; x, z) \in \mathcal{W}$ at a sufficiently fast rate. We therefore restrict our attention to a subset $\mathcal{W}_0$ such that $f(y \mid x, z)$ is bounded away from zero on it.

Based on Corollary 2.2, consider the following null hypothesis

$$\mathbb{H}_0 : \Pr\left[r(Y; X, Z) \circ r_0 - r_1(X) \circ r_2(Y) = 0\right] = 1. \tag{2.5}$$

The alternative hypothesis $\mathbb{H}_A$ is the negation of $\mathbb{H}_0$, i.e.,

$$\mathbb{H}_A : \Pr\left[r(Y; X, Z) \circ r_0 - r_1(X) \circ r_2(Y) = 0\right] < 1. \tag{2.6}$$

According to the characterization result in Theorem 2.1, rejection of (2.5) can only be due to the violation of either $\mathbb{H}_{10}$, the original null hypothesis of interest, or the conditional exogeneity condition in A.1. Maintaining the conditional exogeneity assumption, we may therefore use the null hypothesis $\mathbb{H}_0$ to test the original null of interest, $\mathbb{H}_{10}$. Alternatively, if we maintain the transformation model specification in $\mathbb{H}_{10}$, our test can be used to test the conditional exogeneity assumption A.1 (see remark 2.6 below for more on this last point).

To test the null hypothesis $\mathbb{H}_0$ in (2.5), we use a construction analogous to that of Härdle and Mammen (1993) by considering the weighted L_2 distance between $r \circ r_0$ and $r_1 \circ r_2$:

$$\Gamma \equiv E\left[\left\|r(Y; X, Z) \circ r_0 - r_1(X) \circ r_2(Y)\right\|^2 a(Y; X, Z)\right], \tag{2.7}$$

where $\|\cdot\|$ denotes the Euclidean norm, and $a(y; x, z)$ is a nonnegative weight function that has compact support $\mathcal{Y}_0 \times \mathcal{V}_0$, where $\mathcal{V}_0 \subset \mathcal{V}$. Then $\Gamma = 0$ under $\mathbb{H}_0$ and generally deviates from zero under $\mathbb{H}_A$. In the next subsection we consider the sample version of Γ based on local polynomial estimates of r, r_0, r_1, and r_2.

Remark 2.4. In an previous version of this paper, we considered another set of testable implications:

$$\Pr\left[r(Y; X, Z)\, r_0^{\pi} - r_1(X)\, r_2^{\pi}(Y) = 0\right] = 1 \text{ under } \mathbb{H}_{10}, \tag{2.8}$$

where $r_2^{\pi}(y) \equiv \pi' r_2(y)$, $r_0^{\pi} \equiv \pi' r_0$, and $\pi \equiv (\pi_1, ..., \pi_{d_x})'$ as a $d_x \times 1$ weight vector (e.g., $\pi = (1/d_x, ..., 1/d_x)$). This exploits the additional implication that $s_2(\cdot)$ is a scalar function under $\mathbb{H}_{10}$, however, constructing a test based on equation (2.8) introduces the problem of choosing a weight vector π. Thus, following the suggestion of a referee, we now construct a test baed on the null hypothesis (2.5) above, which is free of π. Note that when $d_x = 1$, π can only equal 1, in which case (2.5) and (2.8) are equivalent.

Remark 2.5. The null $\mathbb{H}_0$ only exploits the implication of Theorem 2.1(a). The rest of this theorem shows that there are some minor additional testable restrictions in $\mathbb{H}_{10}$ that $\mathbb{H}_0$ ignores. First, for technical reasons we have introduced $\mathbf{1}\{y \in \mathcal{Y}_0\}$ to permit trimming of the data as explained in Remark 2.3. So when $\mathcal{Y}_0$ is not equal to the support of Y, our test ignores the trimmed out observations $\{Y_i : Y_i \notin \mathcal{Y}_0\}$. Apart from this trimming, when $d_x = 1$ the only additional restriction implied by $\mathbb{H}_{10}$ that $\mathbb{H}_0$ ignores is a sign restriction on $s_2(\cdot)$, which is $s_2 : \mathbb{R} \to \mathbb{R}_+$ or $s_2 : \mathbb{R} \to \mathbb{R}_-$. That is, the sign of $s_2(y)$ does not depend on y. In theory we could exploit this restriction by testing

$$r(Y; X, Z) \cdot |r_0| - r_1(X) \cdot |r_2(Y)| = 0. \tag{2.9}$$

However, a test based on (2.9) would involve the non-differentiable absolute value function, which would greatly complicate the asymptotic analysis.

Again based on Theorem 2.1, when $d_x > 1$ there are, in addition to the above, two more restrictions in $\mathbb{H}_{10}$ that $\mathbb{H}_0$ ignores, as follows. (i) We do not exploit the fact that $s_2(\cdot)$ is a scalar function under $\mathbb{H}_{10}$. This information could be incorporated into our test as discussed in Remark 2.4. (ii) We do not exploit the fact that $s_1(\cdot)$, a d_x-multivariate function, equals a vector of the derivatives of an unknown scalar function, i.e., $s_1(\cdot)$ can be written as $s_1(x) = -\partial S_1(x)/\partial x$ for some measurable function $S_1 : \mathbb{R}^{d_x} \to \mathbb{R}$. It might be possible to exploit this a restriction in the null hypothesis by imposing the implied constraint that the matrix of derivatives $\partial s_1(x)/\partial x'$ is symmetric (though this would require an additional smoothness assumption). Overall, these differences between $\mathbb{H}_{10}$ and $\mathbb{H}_0$ appear to be relatively minor, so the main substance of the testable implications of $\mathbb{H}_{10}$ is given by the product form $\mathbb{H}_0$ that we test.

Remark 2.6. Chiappori et al. (2013, CKK hereafter) consider a model that is similar to our null model. Using our notation, their model can be written as

$$Y = G\left(\mathcal{H}_1(X, Z) + \tilde{U}\right), \tag{2.10}$$

where X and Z are $d_x \times 1$ and $d_z \times 1$ vectors of observable variables, respectively, $\tilde{U}$ is a scalar unobservable error term, and $G : \mathbb{R} \to \mathbb{R}$ is an unknown strictly monotone function and $\mathcal{H}_1 : \mathbb{R}^{d_x + d_z} \to \mathbb{R}$ is an unknown measurable function. They assume that $X \perp \tilde{U} \mid Z$ and mainly focus on the identification and estimation of $G(\cdot)$ and $\mathcal{H}_1(\cdot)$.[4] In an intermediate step in the proof of their identification result, CKK show that (again using our notation) $r^0(Y; X, Z) = \tilde{s}_1(X, Z) \cdot s_2(Y)$ a.s., where $\tilde{s}_1 : \mathbb{R}^{d_x + d_z} \to \mathbb{R}$ and $s_2 : \mathbb{R} \to \mathbb{R}$ are two measurable functions. This is similar to our characterization of r^0, though CKK use their result in a completely different way. In particular, CKK use their result to show that $\theta(y; x, z)$ is a constant function of (x, z), where $\theta(y; x, z) = S(y; x, z)/E[S(Y; x, z)]$ and $S(y; x, z) = \int_0^y r^0(u; x, z)\, du$. Equivalently,

$$\int \theta(y; x, z)\, w(x, z)\, d(x, z) - \theta(y; x, z) = 0 \;\forall (y, x, z), \tag{2.11}$$

where $w(\cdot)$ is a weight function such that $\int w(x, z)\, d(x, z) = 1$. CKK then propose a weighted L_2-distance-based test to test the conditional exogeneity condition (i.e., $X \perp \tilde{U} \mid Z$) by testing (2.11).

To see the difference between CKK test and our results, recall that we test implications of $Y = G[H_1(X) + H_2(U)]$ against the general alternative $Y = R(X, U)$, given either $X \perp U \mid Z$ or just $X \perp U$ for an unobserved vector U. In contrast, CKK assume $G(\mathcal{H}_1(X, Z) + \tilde{U})$ holds under both null and alternative, and test if $X \perp \tilde{U} \mid Z$, for a scalar $\tilde{U}$.[5]

2.2 Estimation and test statistic

The derivation in the previous section allows the covariates Z to be continuous or discrete. To describe our estimators and associated test statistics, we first consider the (more difficult) case where Z is continuous.

[4] To identify $\mathcal{H}_1(\cdot)$, and hence for estimation, CKK also assume that they observe an additional instrumental variable Q such that $E(\tilde{U} \mid Q) = 0$ and that the conditional distribution of Z given Q is complete.

[5] Other differences are that our test is based directly on Corollary 2.2 while CKK's is based on equation (2.11), we use local polynomials instead of Nadaraya-Watson kernel estimators, and our test is numerically simpler by not requiring any numerical integration, while CKK require multiple numerical integration steps.

Remark 2.7 below then discusses the case where some or all of the elements of Z are discrete.

We employ local polynomial regression to estimate various unknown population objects. Let $v \equiv (x', z')' = (v_1, ..., v_d)'$ be a $d \times 1$ vector, $d \equiv d_x + d_z$, where x is $d_x \times 1$ and z is $d_z \times 1$. Let $\boldsymbol{j} \equiv (j_1, ..., j_d)$ be a d-vector of non-negative integers. Following Masry (1996), adopt the notation

$$v^{\boldsymbol{j}} \equiv \boldsymbol{\Pi}_{i=1}^{d} v_i^{j_i}, \ \boldsymbol{j}! \equiv \boldsymbol{\Pi}_{i=1}^{d} j_i!, \ |\boldsymbol{j}| \equiv \sum_{i=1}^{d} j_i, \quad \sum_{0 \leq |\boldsymbol{j}| \leq p} \equiv \sum_{k=0}^{p} \sum_{j_1=0}^{k} \cdots \sum_{\substack{j_d=0 \\ j_1+\cdots+j_d=k}}^{k} .$$

From $v^{\boldsymbol{j}} \equiv \boldsymbol{\Pi}_{i=1}^{d} v_i^{j_i}$, the j_i's represent powers applied to the elements of v when constructing polynomials.

Consider the p-th order local polynomial estimators $D_x \hat{F}_b(y|x, z)$ of $D_x F(y|x, z)$. The subscript $b = b_n$ is a bandwidth parameter. Let $V_i \equiv (X_i', Z_i')'$ so $V_i - v = ((X_i - x)', (Z_i - z)')'$. Given observations $\{(Y_i, V_i), \ i = 1, ..., n\}$, we estimate $D_x F(y|v)$ by solving the weighted least squares problem

$$\min_{\boldsymbol{\beta}} n^{-1} \sum_{i=1}^{n} \left[\mathbf{1}\{Y_i \leq y\} - \sum_{0 \leq |\boldsymbol{j}| \leq p} \beta_{\boldsymbol{j}}' \left((V_i - v)/b\right)^{\boldsymbol{j}} \right]^2 K_b(V_i - v). \tag{2.12}$$

Here $\boldsymbol{\beta}$ stacks the $\beta_{\boldsymbol{j}}$'s ($0 \leq |\boldsymbol{j}| \leq p$) in lexicographic order (with $\beta_{\mathbf{0}}$, indexed by $\mathbf{0} \equiv (0, ..., 0)$, in the first position, the element with index $(0, 0, ..., 1)$ next, etc.) and $K_b(\cdot) \equiv K(\cdot/b)/b^d$, where $K(\cdot)$ is a symmetric PDF on $\mathbb{R}^d$. Let $\hat{\boldsymbol{\beta}}(y|v)$ denote the solution to the above minimization problem.

Let $N_l \equiv (l + d - 1)!/(l!(d - 1)!)$ be the number of distinct d-tuples $\boldsymbol{j}$ having $|\boldsymbol{j}| = l$. In the above estimation problem, this denotes the number of distinct lth order partial derivatives of $F(y|v)$ with respect to v. Let $N \equiv \sum_{l=0}^{p} N_l$. Let $\mu(\cdot)$ be a stacking function such that $\mu((V_i - v)/b)$ denotes an $N \times 1$ vector that stacks $((V_i - v)/b)^{\boldsymbol{j}}$, $0 \leq |\boldsymbol{j}| \leq p$, in lexicographic order (e.g., $\mu(v) = (1, v')'$ when $p = 1$). Let $\mu_b(v) \equiv \mu(v/b)$. Then

$$\hat{\boldsymbol{\beta}}(y|v) = [\mathbf{S}_b(v)]^{-1} n^{-1} \sum_{i=1}^{n} K_b(V_i - v) \mu_b(V_i - v) \ \mathbf{1}\{Y_i \leq y\}, \tag{2.13}$$

where $\mathbf{S}_b(v) \equiv n^{-1} \sum_{i=1}^{n} K_b(V_i - v) \mu_b(V_i - v) \mu_b(V_i - v)'$. The p-th order local polynomial estimator $D_x \hat{F}_b(y|x, z)$ of $D_x F(y|x, z)$ is given by

$$D_x \hat{F}_b(y|x, z) = e_1 \hat{\boldsymbol{\beta}}(y|x, z)/b \tag{2.14}$$

where $e_1 \equiv [0_{d_x \times 1}, \ I_{d_x}, 0_{d_x \times (N - d_x - 1)}]$ selects the estimator of the coefficient of $(X_i - x)/b$ in the above regression.

To estimate $f(y|v)$, the conditional PDF of Y_i given $V_i = v$, we again employ local polynomial regression. Like Fan et al. (1996), we estimate $f(y|v)$ as $\hat{f}_c(y|v)$, the minimizing constant in the weighted least squares problem

$$\min_{\boldsymbol{\gamma}} n^{-1} \sum_{i=1}^{n} \left[L_c(Y_i - y) - \sum_{0 \leq |\boldsymbol{j}| \leq p} \gamma_{\boldsymbol{j}}' \left((V_i - v)/c\right)^{\boldsymbol{j}} \right]^2 K_c(V_i - v),$$

where $\boldsymbol{\gamma}$ stacks the $\gamma_{\boldsymbol{j}}$'s ($0 \leq |\boldsymbol{j}| \leq p$) in lexicographic order and $L_c(\cdot) \equiv L(\cdot/c)/c$, with $L(\cdot)$ a symmetric kernel function defined on $\mathbb{R}$ and $c \equiv c_n$ a bandwidth parameter. Here, we use the same bandwidth sequence for Y_i and V_i, although different choices of bandwidths are also possible. To reduce the bias of the estimator $\hat{f}_c$, we permit use of a higher-order kernel for L. It is straightforward to verify that

$$\hat{f}_c(y|v) = e_2' [\mathbf{S}_c(v)]^{-1} n^{-1} \sum_{i=1}^{n} K_c(V_i - v) \mu_c(V_i - v) \ L_c(Y_i - y), \tag{2.15}$$

where $e_2 \equiv (1, 0, ..., 0)'$ is an $N \times 1$ vector.[6]

Define

$$\hat{r}(y; x, z) \equiv \frac{D_x \hat{F}_b(y|x,z)}{\hat{f}_c(y|x,z)} \mathbf{1}\{y \in \mathcal{Y}_0\}, \qquad \hat{r}_0 \equiv \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{r}(Y_i; X_j, Z_j),$$
$$\hat{r}_1(x) \equiv \frac{1}{n} \sum_{i=1}^{n} \hat{r}(Y_i; x, Z_i), \quad \text{and} \quad \hat{r}_2(y) \equiv \frac{1}{n} \sum_{i=1}^{n} \hat{r}(y; X_i, Z_i).$$

Our proposed test statistic is

$$\hat{\Gamma} = \frac{1}{n} \sum_{i=1}^{n} \| \hat{r}(Y_i; X_i, Z_i) \circ \hat{r}_0 - \hat{r}_1(X_i) \circ \hat{r}_2(Y_i) \|^2 a(Y_i; X_i, Z_i), \tag{2.16}$$

which is a sample analogue of Γ in (2.7). We next study the asymptotic properties of $\hat{\Gamma}$ under $\mathbb{H}_0$, $\mathbb{H}_A$, and a sequence of Pitman local alternatives.

Remark 2.7. The above estimators and associated tests are easily extended to allow some or all elements of Z to be discrete. To estimate $r(y; x, z)$ in this case, just stratify the sample by each distinct discrete outcome. Specifically, suppose $Z = (Z_c, Z_d)$, where Z_c is continuous and Z_d discrete. Then estimate $r(y; x, z) = r(y; x, z_c, z_d)$ as above (replacing Z with Z_c everywhere), just using the data having $Z_{di} = z_d$, and repeat for each value z_d in the support of Z_d. The functions r_0, r_1, and r_2 can be estimated exactly the same way, by averaging out (X_i, Y_i, Z_i), (Y_i, Z_i), and (X_i, Z_i), respectively, and then our test statistic $\hat{\Gamma}$ is still given by (2.16). More sophisticated estimators (e.g., smoothing across the discrete Z_d cells as proposed in Li and Racine, 2003) could also be used to estimate r these functions. We omit the details for brevity.

3 Asymptotic Properties of the Test Statistic

3.1 Basic assumptions

To study asymptotic properties of $\hat{\Gamma}$, make the following assumptions.

Assumption C.1. Let $W_i \equiv (Y_i, X_i', Z_i')'$, $i = 1, 2, ..., n$, be IID random variables on $(\Omega, \mathcal{F}, P)$, with (Y_i, X_i, Z_i) distributed identically to (Y, X, Z).

Assumption C.2. (i) The PDF $f(v)$ of V_i is continuous in $v \in \mathcal{V}$, and $f(y|v)$ is continuous in $(y, v) \in \mathcal{Y}_0 \times \mathcal{V}$.

(ii) There exist $C_1, C_2 \in (0, \infty)$ such that $C_1 \leq \inf_{v \in \mathcal{V}} f(v) \leq \sup_{v \in \mathcal{V}} f(v) \leq C_2$, and $C_1 \leq \inf_{(y,v) \in \mathcal{Y}_0 \times \mathcal{V}} f(y|v) \leq \sup_{(y,v) \in \mathcal{Y}_0 \times \mathcal{V}} f(y|v) \leq C_2$.

Assumption C.3. (i) $F(\cdot|v)$ is equicontinuous on $\mathcal{Y}_0$: $\forall \epsilon > 0$, $\exists \delta > 0 : |y - \tilde{y}| < \delta \Rightarrow \sup_{y \in \mathcal{Y}_0} |F(y|v) - F(\tilde{y}|v)| < \epsilon$. For each $y \in \mathcal{Y}_0$, $F(y \mid \cdot)$ is Lipschitz continuous on $\mathcal{V}$ and has all partial derivatives up to order $p+1$, $p \in \mathbb{N}$.

(ii) Let $D^{\mathbf{j}} F(y|v) \equiv \partial^{|\mathbf{j}|} F(y|v) / \partial^{j_1} v_1 ... \partial^{j_d} v_d$. For each $y \in \mathcal{Y}_0$, $D^{\boldsymbol{j}} F(y \mid \cdot)$ with $|\boldsymbol{j}| = p+1$ is uniformly bounded and Lipschitz continuous on $\mathcal{V}$: for all $v, \tilde{v} \in \mathcal{V}$, $|D^{\boldsymbol{j}} F(y \mid v) - D^{\boldsymbol{j}} F(y \mid \tilde{v})| \leq C_3 ||v - \tilde{v}||$ for some $C_3 \in (0, \infty)$ where $\|\cdot\|$ is the Euclidean norm.

[6] To ensure that the estimator of $f(y \mid v)$ is positive, we can replace $\hat{f}_c(y \mid v)$ here with a trimmed version defned as $\max\{\hat{f}_c(y \mid v), \epsilon/\sqrt{n}\}$, where ϵ is a small positive number, e.g, $\epsilon = 0.01$. This would not change our resulting asymptotic theory.

(iii) For each $v \in \mathcal{V}$ and for all $y, \tilde{y} \in \mathcal{Y}_0$, $|D^{\boldsymbol{j}} F(y \mid v) - D^{\boldsymbol{j}} F(\tilde{y} \mid v)| \le C_4 \, |y - \tilde{y}|$ for some $C_4 \in (0, \infty)$ where $|\boldsymbol{j}| = p + 1$.

Assumption C.4. Let $r \ge 2$. The rth derivative $f^{(r)}(y|v)$ of $f(y|v)$ with respect to y and all the $(p+1)$th partial derivatives of $f(y|v)$ with respect to v are uniformly continuous on $\mathcal{Y}_0 \times \mathcal{V}$.

Assumption C.5. (i) The kernel $K : \mathbb{R}^d \to \mathbb{R}^+$ is a continuous, bounded, and symmetric PDF.

(ii) $v \to \|v\|^{2p+1} K(v)$ is integrable on $\mathbb{R}^d$ with respect to the Lebesgue measure.

(iii) Let $\mathbf{K}_{\boldsymbol{j}}(u) \equiv v^{\boldsymbol{j}} K(v)$ for all $\boldsymbol{j}$ with $0 \le |\boldsymbol{j}| \le 2p+1$. For some finite constants σ_K, $\bar{\sigma}_1$, and $\bar{\sigma}_2$, either $K(\cdot)$ is compactly supported such that $K(v) = 0$ for $\|u\| > \sigma_K$, and $|\mathbf{K}_{\boldsymbol{j}}(v) - \mathbf{K}_{\boldsymbol{j}}(\tilde{v})| \le \bar{\sigma}_2 \|v - \tilde{v}\|$ for any $v, \tilde{v} \in \mathbb{R}^d$ and for all $\boldsymbol{j}$ with $0 \le |\boldsymbol{j}| \le 2p+1$; or $K(\cdot)$ is differentiable, $\|\partial \mathbf{K}_{\boldsymbol{j}}(v) / \partial v\| \le \bar{\sigma}_1$, and for some $\iota_0 > 1$, $|\partial \mathbf{K}_{\boldsymbol{j}}(v) / \partial v| \le \bar{\sigma}_1 \|v\|^{-\iota_0}$ for all $\|v\| > \sigma_K$ and for all $\boldsymbol{j}$ with $0 \le |\boldsymbol{j}| \le 2p+1$.

Assumption C.6. The univariate kernel function L satisfies $\int L(y)^2 dy < \infty$ and is a symmetric rth order kernel, i.e., $\int L(y) dy = 1$, $\int y^s L(y) dy = 0$ for all $s = 1, ..., r-1$, and $\int y^r L(y) dy < \infty$. The rth derivative of L exists and is continuous.

Assumption C.7. (i) $p > d/2$.

(ii) As $n \to \infty$, $(c^{p+1} + c^r)/b^{d/2} \to 0$, $(b^p + c^{p+1} + c^r) b^{d/2+2}/c^{d+1} \to 0$, $b^{d+4}/c^{d+1} \to 0$, $nb^{2(p+1)+d} \to 0$, and $nb^{d+2}(c^{2(p+1)} + c^{2r}) \to 0$.

(iii) As $n \to \infty$, $\min\{nb^{2d},\ nb^{3d/2+1}/\ln n,\ nb^{d+2},\ nb^{d_x+2}/\ln n,\ nb^{d+1}c^{(d+1)/2}/\ln n,\ nb^{d/2}c^{d+1}/\ln n,$ $nb^{-(d/2+2)}c^{2(d+1)}/\ln n,\ nb^{-1}c^{3(d+1)/2}/\ln n,\ nb^{-(d+4)}c^{3(d+1)}\} \to \infty$.

We assume IID observations in Assumption C.1, which is standard in cross-section studies. Assumptions C.2-C.4 impose smoothness conditions on the conditional CDF $(y|v)$ and PDF $f(y|v)$ that are used to ensure uniform consistency of our local polynomial estimators, based on results of Masry (1996) and Hansen (2008). Assumptions C.5 and C.6 impose conditions on the kernels K and L, which are standard in the literature for local polynomial regression or conditional density estimation. Assumption C.7 restricts the choice of bandwidth sequences b and c, the order p of local polynomial regressions, and the order r of the kernel L. This assumption allows c to differ from b, but in the case where $b = c$ Assumption C.7 simplifies to the following assumption:

Assumption C.7*. (i) $p > d/2$ and $r > d/2$.

(ii) As $n \to \infty$, $nb^{2(p+1)+d} \to 0$ and $nb^{2r+d+2} \to 0$.

(iii) As $n \to \infty$, $\min\{nb^{2d},\ nb^{3(d+1)/2}/\ln n,\ nb^{d+2},\ nb^{d_x+2}/\ln n\} \to \infty$.

Note that we allow $d_z = 0$, otherwise the condition $nb^{d_x+2}/\ln n \to \infty$ as $n \to \infty$ becomes redundant.

3.2 Asymptotic null distribution

In this section, we study the asymptotic behavior of the test statistic in (2.16). To state the next result, let $w_i \equiv (y_i, v_i')'$ and introduce the following notation:

$$
\begin{aligned}
\zeta_{1k}(y; v) &\equiv b^{-1} f(y|v)^{-1} e_1 \bar{\mathbf{S}}_b(v)^{-1} \mu_b(V_k - v) K_b(V_k - v) \bar{\mathbf{I}}_y(W_k) \mathbf{1}\{y \in \mathcal{Y}_0\}, \\
\zeta_{2k}(y; v) &\equiv f(y|v)^{-2} D_x F(y|v) e_2' \bar{\mathbf{S}}_c(v)^{-1} \mu_c(V_k - v) K_c(V_k - v) \bar{\mathbf{L}}_y(W_k) \mathbf{1}\{y \in \mathcal{Y}_0\}, \\
\zeta_k(y; v) &\equiv \zeta_{1k}(y; v) - \zeta_{2k}(y; v), \\
\zeta(W_i, W_j, W_k) &\equiv \left(\zeta_j(Y_i; V_i) \circ r_0\right)' \left(\zeta_k(Y_i; V_i) \circ r_0\right) a_i, \\
\varphi(w_i, w_j) &\equiv E\left[\zeta(W_1, w_i, w_j)\right],
\end{aligned}
$$

where $\bar{\mathbf{S}}_b(v) \equiv E[\mathbf{S}_b(v)]$, $\bar{\mathbf{I}}_y(W_k) \equiv \mathbf{1}\{Y_k \leq y\} - F(y|V_k)$, $\bar{\mathbf{L}}_y(W_k) \equiv L_c(Y_k - y) - \alpha(y|V_k)$, $\alpha(y|v) \equiv E[L_c(Y_k - y)|V_k = v]$ and $a_i \equiv a(Y_i; X_i, Z_i)$. Define the asymptotic bias term

$$
\begin{aligned}
\mathbb{B}_n &\equiv n^{-1} b^{\frac{d}{2}+2} \sum_{i=1}^{n} \varphi(W_i, W_i) + n^{-4} b^{\frac{d}{2}+2} \sum_{i=1}^{n} \left\| \sum_{j=1}^{n} \sum_{k=1}^{n} \zeta_k(Y_j; X_i, Z_j) \circ r_2(Y_i) \right\|^2 a_i \\
&\quad -2n^{-3} b^{\frac{d}{2}+2} \sum_{i=1}^{n} \left(\sum_{l=1}^{n} \zeta_l(Y_i; V_i) \circ r_0 \right)' \left(\sum_{j=1}^{n} \sum_{k=1}^{n} \zeta_k(Y_j; X_i, Z_j) \circ r_2(Y_i) \right) a_i \\
&\equiv \mathbb{B}_{1n} + \mathbb{B}_{2n} - 2\mathbb{B}_{3n}, \text{ say.}
\end{aligned}
$$

We establish the asymptotic null distribution of the $\hat{\Gamma}$ test statistic as follows:

Theorem 3.1 *Suppose Assumptions A.1 and C.1-C.7 hold. Then*

$$
nb^{\frac{d}{2}+2}\hat{\Gamma} - \mathbb{B}_n \xrightarrow{d} N\left(0, \sigma_0^2\right),
$$

where $\sigma_0^2 \equiv \lim_{n\to\infty} \sigma_n^2$ *and* $\sigma_n^2 = 2b^{d+4} E[\varphi(W_1, W_2)^2]$.

Remark 3.1. The asymptotic bias $\mathbb{B}_n$ of $nb^{\frac{d}{2}+2}\hat{\Gamma}$ contains three terms $\mathbb{B}_{1n}$, $\mathbb{B}_{2n}$, and $-2\mathbb{B}_{3n}$. The first two terms reflect the contributions of $\hat{r}(Y_i; V_i) \circ \hat{r}_0$ and $\hat{r}_1(X_i) \circ \hat{r}_2(Y_i)$, respectively, and the last term reflects the interaction between these latter two terms. We show that $\mathbb{B}_{1n} = O_P(b^{\frac{d}{2}+2}(b^{-d-2} + c^{-d-1}))$ in Lemma B.4, $\mathbb{B}_{2n} = O_P(b^{\frac{d}{2}+2}(b^{-d_x-2} + c^{-d_x}))$ in Lemma B.5(b), and $\mathbb{B}_{3n} = O_P(b^{\frac{d}{2}+2}(b^{-d_x-2} + c^{-d_x}))$ in Lemma B.6(b). Here $\mathbb{B}_{1n}$ never vanishes asymptotically whereas $\mathbb{B}_{2n}$ and $\mathbb{B}_{3n}$ are asymptotically negligible under appropriate conditions, such as when $b = c$ and $d_z > d_x$. The asymptotic variance σ_n^2 of $nb^{\frac{d}{2}+2}\hat{\Gamma}$ only reflects the contribution of $\hat{r}(Y_i; V_i) \circ \hat{r}_0$, due to the faster convergence rate of $\hat{r}_1(X_i) \circ \hat{r}_2(Y_i)$ to $r_1(X_i) \circ r_2(Y_i)$ than that of $\hat{r}(Y_i; V_i) \circ \hat{r}_0$ to $r(Y_i; V_i) \circ r_0$.

To implement the test, we need consistent estimates of the asymptotic bias and variance. Let

$$
\begin{aligned}
\hat{\zeta}_{1k}(y; v) &\equiv b^{-1} \hat{f}_c(y|v)^{-1} e_1 \mathbf{S}_b(v)^{-1} \mu_b(V_k - v) K_b(V_k - v) \hat{\mathbf{I}}_y(W_k) \mathbf{1}\{y \in \mathcal{Y}_0\}, \\
\hat{\zeta}_{2k}(y; v) &\equiv \hat{f}_c(y|v)^{-2} D_x \hat{F}_b(y|v) e_2' \mathbf{S}_c(v)^{-1} \mu_c(V_k - v) K_c(V_k - v) \hat{\mathbf{L}}_y(W_k) \mathbf{1}\{y \in \mathcal{Y}_0\}, \\
\hat{\zeta}_k(y; v) &\equiv \hat{\zeta}_{1k}(y; v) - \hat{\zeta}_{1k}(y; v), \\
\hat{\varphi}(W_j, W_k) &\equiv n^{-1} \sum_{i=1}^{n} \left(\hat{\zeta}_j(Y_i; V_i) \circ \hat{r}_0 \right)' \left(\hat{\zeta}_k(Y_i; V_i) \circ \hat{r}_0 \right) a_i,
\end{aligned}
$$

where $\hat{\mathbf{I}}_y(W_k) \equiv \mathbf{1}\{Y_k \leq y\} - \hat{F}_b(y|V_k)$, $\hat{\mathbf{L}}_y(W_k) \equiv L_c(Y_k - y) - \hat{f}_c(y|V_k)$, and $\hat{F}_b(y|V_k)$ is the pth order local polynomial estimator of $F(y|V_k)$ by using the kernel K and bandwidth b. We propose to estimate the asymptotic bias $\mathbb{B}_n$ and variance σ_n^2 respectively by

$$
\begin{aligned}
\hat{\mathbb{B}}_n &\equiv n^{-1} b^{\frac{d}{2}+2} \sum_{i=1}^{n} \hat{\varphi}(W_i, W_i) + n^{-4} b^{\frac{d}{2}+2} \sum_{i=1}^{n} \left\| \sum_{j=1}^{n} \sum_{k=1}^{n} \hat{\zeta}_k(Y_j; X_i, Z_j) \circ \hat{r}_2(Y_i) \right\|^2 a_i \\
&\quad -2n^{-3} b^{\frac{d}{2}+2} \sum_{i=1}^{n} \left(\sum_{l=1}^{n} \hat{\zeta}_l(Y_i; V_i) \circ \hat{r}_0 \right)' \left(\sum_{j=1}^{n} \sum_{k=1}^{n} \hat{\zeta}_k(Y_j; X_i, Z_j) \circ \hat{r}_2(Y_i) \right) a_i, \\
\hat{\sigma}_n^2 &\equiv 2n^{-2} b^{d+4} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{\varphi}(W_i, W_j)^2.
\end{aligned}
$$

It is straightforward to show $\hat{\mathbb{B}}_n - \mathbb{B}_n = o_P(1)$ and $\hat{\sigma}_n^2 - \sigma_n^2 = o_P(1)$. We can now compare

$$T_n \equiv \left(nb^{\frac{d}{2}+2}\hat{\Gamma} - \hat{\mathbb{B}}_n\right)/\sqrt{\hat{\sigma}_n^2} \tag{3.1}$$

to the critical value z_α defined as the upper α percentile from the $N(0,1)$ distribution (since the test is one-sided) and reject the null when $T_n > z_\alpha$.

3.3 Consistency and asymptotic local power

The following theorem shows that the test T_n is consistent for the class of global alternatives

$$\mathbb{H}_A : \mu_A \equiv E\left\{\|r(Y;X,Z) \circ r_0 - r_1(X) \circ r_2(Y)\|^2 a(Y;X,Z)\right\} > 0.$$

Theorem 3.2 *Suppose Assumptions C.1-C.7 hold. Then T_n diverges to infinity at the rate of $nb^{\frac{d}{2}+2}$ under $\mathbb{H}_A$, i.e., $P(T_n > e_n) \to 1$ as $n \to \infty$ under $\mathbb{H}_A$ for any nonstochastic sequence $e_n = o(nb^{\frac{d}{2}+2})$.*

To study the local power of our test, we focus on the case $H_2(U) = U$ (which requires $d_u = 1$) and consider the model

$$Y_{ni} \equiv R_n(X_{ni}, U_{ni}) = G[H_1(X_{ni}) + U_{ni} + \gamma_n \delta(X_{ni}, U_{ni})] \tag{3.2}$$

where $\gamma_n \to 0$ as $n \to \infty$, $\delta(X_{ni}, U_{ni})$ is not additively separable in X_{ni} and U_{ni}, and G and H_1 are as defined under $\mathbb{H}_{10}$ in Section 1. Note that we allow both Y_{ni} and $(X_{ni,}U_{ni})$ to be double-array and the structural function R_n is now n-dependent. As before, we assume that X_{ni} and U_{ni} are independent given the d_z-vector $Z_{ni} : X_{ni} \perp U_{ni} \mid Z_{ni}$. Following the literature on nonseparable models, we also assume that $R_n(x, \cdot)$ is strictly increasing for each x on the support of X_{ni}. Formally, we put these requirements into the following assumption.

Assumption A.1* $X_{ni} \perp U_{ni} \mid Z_{ni}$ and equation (3.2) holds such that $R_n(x, \cdot)$ is strictly increasing for each x on the support of X_{ni}.

Let $\delta_n(x,u) \equiv u + \gamma_n \delta(x,u)$. Given the strict monotonicity of G, without loss of generality, we assume G is also strictly increasing. This, in conjunction with Assumption A.1* implies that $\delta_n(x, \cdot)$ is strictly increasing for each x on the support of X_{ni}.

Let $F_n(\cdot \mid x, z)$ and $f_n(\cdot \mid x, z)$ denote the conditional CDF and PDF of Y_{ni} given $(X_{ni}, Z_{ni}) = (x, z)$, respectively. We now use $\mathcal{X}_n$, $\mathcal{Z}_n$, and $\mathcal{W}_n$ to denote the supports of X_{ni}, Z_{ni}, and (Y_{ni}, X_{ni}, Z_{ni}), respectively. Let $r_n^0(y;x,z) \equiv \frac{D_x F_n(y|x,z)}{f_n(y|x,z)}$. The following theorem parallels Theorem 2.1(a) and lays down the foundation for the asymptotic local analysis of our test statistic.

Theorem 3.3 *Suppose that A.1* holds. Suppose that $f_n(y \mid x, z)$ is continuously differentiable with respect to both y and x for each $z \in \mathcal{Z}_n$ and $f_n(y \mid x,z) \neq 0$ for all $(y,x,z) \in \mathcal{W}_n$. Suppose that $G : \mathbb{R} \to \mathbb{R}$ is strictly increasing and continuously differentiable, $H_1 : \mathbb{R}^{d_x} \to \mathbb{R}$ is continuously differentiable, and $\delta(x, \cdot)$ is continuously differentiable for each $x \in \mathcal{X}_n$. Then there exist two measurable functions $s_{1n} : \mathbb{R}^{d_x} \to \mathbb{R}^{d_x}$ and $s_2 : \mathbb{R} \to \mathbb{R}_+$ or $s_2 : \mathbb{R} \to \mathbb{R}_-$ such that*

$$r_n^0(y;x,z) = s_{1n}(x)\, s_2(y) + \gamma_n \Delta_n(y;x,z) + o(\gamma_n) \ \text{ for all } \ (y,x,z) \in \mathcal{W}_n \tag{3.3}$$

where $s_{1n}(x) = -\partial S_{1n}(x)/\partial x$ for some measurable function $S_{1n} : \mathbb{R}^{d_x} \to \mathbb{R}$, $1/s_2(y) = \partial S_2(y)/\partial y$ for some measurable function $S_2 : \mathbb{R} \to \mathbb{R}$, and $\Delta_n(y;x,z)$ is some measurable function of (y,x,z).

As in Section 2.1, let $r_n(y;x,z) \equiv r_n^0(y;x,z)\mathbf{1}\{y \in \mathcal{Y}_0\}$, $r_{0n} \equiv E_{Y_n}E_{X_nZ_n}[r_n(Y_{ni};X_{ni},Z_{ni})]$, $r_{1n}(x) \equiv E[r_n(Y_{ni};x,Z_{ni})]$, and $r_{2n}(y) \equiv E[r_n(y;X_{ni},Z_{ni})]$. The following corollary indicates that $r_{1n}(x) \circ r_{2n}(y)$ only deviates from $r_n(y;x,z) \circ r_{0n}$ locally.

Corollary 3.4 *Suppose that the conditions in Theorem 3.3 hold. Then*

$$r_n(y;x,z) \circ r_{0n} - r_{1n}(x) \circ r_{2n}(y) = \gamma_n \bar{\Delta}_n(y;x,z) + o(\gamma_n) \text{ for all } (y,x,z) \in \mathcal{W}_{0n},$$

where $\bar{\Delta}_n(y;x,z)$ is defined in (B.6) in the appendix and $\mathcal{W}_{0n}$ is defined as $\mathcal{W}_n$ but with y restricted on $\mathcal{Y}_0$.

With the above corollary, we can study the local power of our test. We consider the following sequence of Pitman local alternatives:

$$\mathbb{H}_A(\gamma_n): r_n(y;x,z) \circ r_{0n} - r_{1n}(x) \circ r_{2n}(y) = \gamma_n \bar{\Delta}_n(y;x,z) + o(\gamma_n), \tag{3.4}$$

where $\gamma_n \to 0$ as $n \to \infty$, and $\bar{\Delta}_n$ is a nonzero measurable function with $\mu_0 \equiv \lim_{n\to\infty} E[\bar{\Delta}_n(Y_{n1};X_{n1},Z_{n1})^2 a(Y_{n1};X_{n1},Z_{n1})] < \infty$. For technical reasons, we assume that the term $o(\gamma_n)$ in (3.4) holds uniformly in (y,x,z) on the support of the weight function $a(y;x,z)$.

We continue to use $\hat{r}$, $\hat{r}_0$, $\hat{r}_1$, and $\hat{r}_2$ to denote the local polynomial estimates of r_n, r_{0n}, r_{1n}, and r_{2n}, respectively. The asymptotic bias and variance terms are estimated as before with slight notational changes to account for double-array processes. The final test statistic T_n is constructed as before. The following theorem reports the asymptotic property of T_n under $\mathbb{H}_A(\gamma_n)$.

Theorem 3.5 *Suppose Assumptions C.1-C.7 hold with the obvious notational changes that allows double-array IID processes. Then under $\mathbb{H}_A(\gamma_n)$ with $\gamma_n = n^{-1/2}b^{-d/4-1}$, $T_n \xrightarrow{d} N(\mu_0/\sigma_0, 1)$.*

Remark 3.2. Theorem 3.5 implies that the T_n test has non-trivial power against Pitman local alternatives that converge to zero at rate $n^{-1/2}b^{-d/4-1}$, provided $0 < \mu_0 < \infty$. The asymptotic local power function of the test is given by $1 - \Phi(z_\alpha - \mu_0/\sigma_0)$, where Φ is the standard normal CDF. It is worth mentioning that the local alternative in (3.4) may be motivated from models other than that considered in (3.2). Generally speaking, the local alternative in (3.4) represents a class of local deviations from the implied null hypothesis $\mathbb{H}_0$ which our test has power to detect. Such a local deviation may be caused by the violation of any conditions specified under $\mathbb{H}_{01}$. This includes the model in (3.2) where $\delta_n(x,u) \equiv u + \gamma_n \delta(x,u)$ may or may not be monotone in its second argument (even though we assume monotonicity to facilitate the derivation), and the case where the conditional exogeneity condition in Assumption A.1* is locally violated.[7]

Remark 3.3. Alternatively, following Remark 2.4, one can exploit the testable implication in (2.8) and construct the test statistic:

$$\hat{\Gamma}(\pi) = \frac{1}{n}\sum_{i=1}^{n} \left\| \hat{r}(Y_i;X_i,Z_i)\hat{r}_0^\pi - \hat{r}_1(X_i)\hat{r}_2^\pi(Y_i) \right\|^2 a(Y_i;X_i,Z_i),$$

where $\hat{r}_0^\pi \equiv \pi'\hat{r}_0$ and $\hat{r}_2^\pi(y) \equiv \pi'\hat{r}_2(y)$. In an earlier version of this paper, we showed that a suitable normalization of $\hat{\Gamma}(\pi)$, say, $T_n(\pi)$, has the standard normal limiting null distribution. This test would

[7] This means the conditional PDF $f_n(\cdot|x)$ of Y_{ni} given $X_{ni} = x$ deviates from the conditional PDF $f_n(\cdot|x,z)$ of Y_{ni} given $X_{ni} = x$ and $Z_{ni} = z$ locally: $f_n(y|x,z) - f_n(y|x) = \gamma_n \eta(y;x,z) + o(\gamma_n)$, where $\gamma_n \to 0$ as $n \to \infty$ and $\eta(y;x,z)$ is a measurable function of (y,x,z).

require selection of a weigh vector π. For convenience, one could choose equal weights $\pi = (1/d_x, ..., 1/d_x)$. A referee suggested that one might improve power by maximizing over possible weight vectors, e.g., basing a test on the statistic $\sup_{\pi \in \mathcal{S}} \hat{\Gamma}(\pi)$ where $\mathcal{S}$ is a unit sphere in $\mathbb{R}^{d_x}$. Such a test would likely be considerably more demanding computationally than our proposed test.

Remark 3.4. Like many nonparametric specification tests in the literature (e.g., Härdle and Mammen (1993), HSW, and CKK), our test suffers from a typical curse of dimensionality. Our test can detect local alternatives that converge to the null at the rate of $\gamma_n = n^{-1/2} b^{-d/4-1}$, so as d increases, the local power of our test deteriorates, and does so at rates that are common for nonparametric tests. In our simulations and empirical applications, d is small. equalling one or two. To alleviate the curse of dimensionality in problems with larger values of d and limited sample sizes, one could consider semiparametric models of $H_1(x)$ under the null hypothesis. Examples of such specifications for $H_1(x)$ could include partially linear models, single-index models, or additive models. See, e.g., Fan et al. (2001) or Fan and Jiang (2005), among others.

3.4 Simulating the null distribution

It is well known that the asymptotic normal null distributions with estimated variance matrices often do not provide good approximations for kernel-based tests, in part because tests based on normal critical values can be very sensitive to the choice of bandwidth and suffer from substantial finite sample size distortions. We found that to be the case in some experiments with our test (not reported to save space). An an alternative, we consider resampling methods to obtain the simulated p-values or critical values for our test.

As discussed earlier in remark 2.6, CKK propose a test related to ours. They employ a nonparametric bootstrap method to obtain p-values for their test, but they do not formally demonstrate its asymptotic validity for technical reasons[8]. In contrast, the asymptotic validity of subsampling can be justified by standard arguments, so we use subsampling instead of a standard bootstrap. Let $m = m_n$ be a sequence of positive integers such that $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. Let B be a large integer. The subsampling (or equivalently, the m-out-of-n bootstrap) procedure is as follows:

1. Randomly draw B subsamples $\left\{\left(X_i^{*(k)}, Y_i^{*(k)}, Z_i^{*(k)}\right), i = 1, ..., m\right\}_{k=1}^{B}$ of size m from the original sample $\{(X_i, Y_i, Z_i)\}_{i=1}^{n}$.
2. For $k = 1, ..., B$, compute T_n using the subsample $\left\{\left(X_i^{*(k)}, Y_i^{*(k)}, Z_i^{*(k)}\right)\right\}_{i=1}^{m}$ and denote this as $\hat{T}_{n,m}^{*(k)}$.
3. Calculate the subsampling p-value as

$$p = B^{-1} \sum_{k=1}^{B} \mathbf{1}\left\{T_n < \hat{T}_{n,m}^{*(k)}\right\}.$$

The asymptotic validity of the above subsampling method can be readily established as in Politis et al. (1999). Intuitively, under the null hypothesis both T_n and $\hat{T}_{n,m}^{*(k)}$ are asymptotically distributed as $N(0,1)$ and thus the test based on the subsampling-based p-value has the correct asymptotic size, and under the fixed alternative T_n diverges to infinity at a speed faster than $\hat{T}_{n,m}^{*(k)}$, giving the test its power.

[8] This bootstrap depends on rates of uniform convergence of kernel estimated objects, and such rates can fail due to issues associated with unbounded support or to boundary biases.

4 Monte Carlo Simulations

In this section, we use simulations to examine the finite sample performance of our test. We consider four data generating processes (DGPs):
DGPs 1 and 3: $Y = X + U + \lambda X\sqrt{1+U^2}$;
DGPs 2 and 4: $Y = \Phi(X + U + \lambda X\sqrt{1+U^2})$;

For each DGP, λ takes four values: 0, 0.25, 0.5, and 1. $\lambda = 0$ corresponds to the null DGP and $\lambda = 0.25$, 0.5 and 1 represent gradual departure from the null. In DGPs 1 and 2, $X \sim$Uniform$(-1,1)$, $U \sim$Uniform$(-1,1)$, and X and U are independent. In DGPs 3 and 4, X and U are no longer independent: $X = 0.5Z + 0.5\varepsilon_1$ and $U = 0.5Z + 0.5\varepsilon_2$, where $\varepsilon_1 \sim$Uniform$(-1,1)$, $\varepsilon_2 \sim$Uniform$(-1,1)$, Z follows a standard normal distribution truncated by -2 and 2 in the tails, and ε_1, ε_2, and Z are mutually independent. By construction, $X \perp U \mid Z$.

We use second order (quadratic) local polynomial estimators, i.e., $p = 2$, with a Gaussian PDF for the kernel function. For the bandwidth sequence b and c, we use the rule $\kappa{\cdot}\mathrm{std}(V) \cdot n^{-\frac{1}{2(p+1)+1}}$ and $\kappa{\cdot}\mathrm{std}(Y) \cdot n^{-\frac{1}{2(p+1)+1}}$ associated with V and Y, respectively, where κ is a constant and std(V) and std(Y) are sample standard deviations of V and Y, respectively. In general, the optimal κ depends on the underlying specific DGPs. For simplicity we let $\kappa = 1$ for DGPs 1 and 2 and $\kappa = 2$ for DGPs 3 and 4. For DGPs 1 and 2, we specify the weight function $a = 1$, corresponding to no trimming, whereas for DGPs 3 and 4, a trims out 2.5% data on each tail of each dimension of (Y, X, Z), so

$$a\,(Y; X, Z) = \mathbf{1}\left[y_{0.025} \le Y \le y_{0.975}\right] \cdot \mathbf{1}\left[x_{0.025} \le X \le x_{0.975}\right] \cdot \mathbf{1}\left[z_{0.025} \le Z \le z_{0.975}\right],$$

where $y_{0.025}$ and $y_{0.975}$ are the 0.025 and 0.975 quantiles of Y respectively, and similarly for $x_{0.025}$, $x_{0.975}$, $z_{0.025}$, and $z_{0.975}$.

We consider the subsampling test with the sample sizes $n = 200$ and 300. We try three different subsample sizes $m = \lfloor n^{0.80} \rfloor$, $\lfloor n^{0.85} \rfloor$, and $\lfloor n^{0.90} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part of $\cdot$. The number of subsamples is $B = 200$ and the number of replications is 500.

We consider two conventional nominal levels: 0.05 and 0.10. Tables 1-4 present the rejection frequencies for DGPs 1-4, respectively. In each Table, $\lambda = 0$ corresponds to the null DGP. When the sample size is 200, the subsampling tests are undersized. However, when the sample size increases to 300, the performance improves and the rejection frequencies are closer to their nominal levels. This suggests that a moderate to large sample is required for the test to have good level behavior. This is not surprising, as the estimation of derivatives is much harder and has a slower convergence rate than the estimation of the conditional expectation itself. In general, the tests are less under-sized when the subsample size m is relatively small.

Table 1: Empirical rejection frequency: DGP 1

λ	n	Subsample size					
		$\lfloor n^{0.80} \rfloor$		$\lfloor n^{0.85} \rfloor$		$\lfloor n^{0.90} \rfloor$	
		0.05	0.10	0.05	0.10	0.05	0.10
0	200	0.006	0.032	0.006	0.038	0.004	0.026
	300	0.034	0.134	0.032	0.134	0.016	0.090
0.25	200	0.124	0.286	0.110	0.258	0.056	0.196
	300	0.428	0.642	0.358	0.653	0.230	0.504
0.5	200	0.460	0.686	0.388	0.622	0.258	0.498
	300	0.860	0.954	0.820	0.926	0.638	0.860
1	200	0.910	0.990	0.864	0.968	0.748	0.908
	300	0.998	1.000	0.996	1.000	0.982	0.996

Table 2: Empirical rejection frequency: DGP 2

λ	n	Subsample size					
		$\lfloor n^{0.80} \rfloor$		$\lfloor n^{0.85} \rfloor$		$\lfloor n^{0.90} \rfloor$	
		0.05	0.10	0.05	0.10	0.05	0.10
0	200	0.010	0.040	0.008	0.040	0.004	0.022
	300	0.036	0.098	0.028	0.092	0.020	0.066
0.25	200	0.186	0.380	0.134	0.344	0.046	0.216
	300	0.412	0.686	0.326	0.592	0.178	0.444
0.5	200	0.700	0.872	0.596	0.836	0.390	0.674
	300	0.938	0.986	0.892	0.970	0.694	0.924
1	200	0.958	0.996	0.878	0.988	0.594	0.908
	300	0.994	1.000	0.980	0.998	0.734	0.972

In each Table, $\lambda = 0.25$, 0.5 and 1 correspond to alternative DGPs and thus they are used to examine the power of the tests. The test has substantial power. For example, at the 5% level, the rejection frequency is 0.910 for DGP 1 when $\lambda = 1$, sample size $n = 200$, and subsample size $m = \lfloor n^{0.80} \rfloor$. For all the values of $\lambda = 0.25$, 0.5 and 1, the power increases rapidly as the sample size increases. For example, when n increases from 200 to 300, the rejection frequency increases from 0.460 to 0.860 for DGP 1 with $\lambda = 0.5$ and $m = \lfloor n^{0.80} \rfloor$. As expected, the rejection frequencies increase with the value of λ for all DGPs. The power of the tests increases when the subsample size m decreases. This is likely because, under the alternative, the test statistics diverge, so the difference between the original test statistics and subsampled test statistics is large when the difference between the original sample size and subsample size is large.

In general, for the same sample size, the rejection frequencies for DGPs 1 and 2 under the alternative are higher than those for DGPs 3 and 4. This suggests that when d is large, we need to have a relatively large sample to achieve reasonable powers. This reflects the "curse of dimensionality" of our test.

Table 3: Empirical rejection frequency: DGP 3

λ	n	Subsample size					
		$\lfloor n^{0.80} \rfloor$		$\lfloor n^{0.85} \rfloor$		$\lfloor n^{0.90} \rfloor$	
		0.05	0.10	0.05	0.10	0.05	0.10
0	200	0.010	0.040	0.000	0.016	0.002	0.014
	300	0.034	0.106	0.018	0.074	0.004	0.034
0.25	200	0.056	0.162	0.034	0.092	0.016	0.074
	300	0.174	0.370	0.122	0.334	0.034	0.174
0.5	200	0.216	0.482	0.122	0.368	0.072	0.242
	300	0.546	0.754	0.446	0.720	0.220	0.514
1	200	0.806	0.928	0.642	0.864	0.484	0.782
	300	0.972	0.990	0.940	0.986	0.808	0.952

Table 4: Empirical rejection frequency: DGP 4

λ	n	Subsample size					
		$\lfloor n^{0.80} \rfloor$		$\lfloor n^{0.85} \rfloor$		$\lfloor n^{0.90} \rfloor$	
		0.05	0.10	0.05	0.10	0.05	0.10
0	200	0.018	0.054	0.012	0.030	0.002	0.022
	300	0.040	0.096	0.022	0.074	0.004	0.044
0.25	200	0.030	0.096	0.020	0.062	0.000	0.038
	300	0.084	0.158	0.048	0.134	0.012	0.074
0.5	200	0.074	0.182	0.038	0.102	0.010	0.068
	300	0.134	0.292	0.098	0.252	0.042	0.142
1	200	0.250	0.482	0.138	0.344	0.060	0.226
	300	0.448	0.678	0.346	0.612	0.166	0.404

5 Empirical Applications

In this section, we consider testing whether duration data obey the class of nonlinear generalized accelerated failure-time (GAFT) models. We then apply our test empirically on a data set of duration of strikes among manufacturing workers in the US.

5.1 Testing for GAFT models

Let Y be the duration of a certain state (a nonnegative random variable) such as duration of a strike. Our test is directly applicable to nonlinear GAFT models, since such models can be written in the form $Y = G\left[H_1(X) + U\right]$, where X is a vector of covariates, and U an unobservable random variable (see, e.g., equation (2.5) in Ridder, 1990).

MPH (Mixed Proportional Hazard) models are a particularly popular class of GAFT models. Below we provide a direct link between our null hypothesis and MPH models. Let $h(Y, X, \xi)$ denote the hazard function for Y. An MPH model of survival time Y is one where

$$h(Y, X, \xi) = \lambda(Y) \cdot \theta(X) \cdot \xi \tag{5.1}$$

holds for some baseline hazard function $\lambda(Y)$ and some nonnegative function of covariates $\theta(X)$. The MPH model is widely applied in empirical research (for a detailed review, see Van den Berg (2001)). For example, when $\xi = 1$, this is the standard proportion hazard (PH) model developed by Cox (1972). A particularly popular parametric specification of the MPH model due to Lancaster (1979) assumes that $\lambda(Y) = \alpha Y^{\alpha-1}$, $\theta(X) = \exp(X'\beta)$ and ξ is a gamma distributed random variable. The following Proposition provides a general characterization of MPH models.

Proposition 5.1 *Suppose that the hazard function of the survival time Y is $h(Y, X, \xi)$, where $Y \in \mathbb{R}_+$, $X \in \mathbb{R}^{d_x}, \xi \in \mathbb{R}_+$ and $\xi \neq 0$ with probability 1. Let $\lambda : \mathbb{R} \to \mathbb{R}_+$ and $\theta : \mathbb{R}^{d_x} \to \mathbb{R}_+$ be two measurable functions such that $\lambda(Y) = 0$ with probability 0 and $\theta(X) = 0$ with probability 0. Then $h(Y, X, \xi)$ is a MPH model:*

$$h(Y, X, \xi) = \lambda(Y) \cdot \theta(X) \cdot \xi,$$

if and only if

$$Y = G[H_1(X) + U],$$

where $G : \mathbb{R} \to \mathbb{R}_+$ is a strictly increasing function that is differentiable a.e. on its support, $H_1 : \mathbb{R}^{d_x} \to \mathbb{R}$, and $U = \ln\left[\frac{-\ln(1-\varepsilon)}{\xi}\right]$, where ε is a uniform random variable on $[0, 1]$ and $\varepsilon \perp (X, \xi)$.

Proposition 5.1 shows that the MPH model has two important implications: (i) it equals a transformation model of the type given by our null, and (ii) U allows a distribution determined by $\ln(-\ln(1-\varepsilon)/\xi)$. In principle, both restrictions might be testable, though we focus on implication (i), corresponding to our null hypothesis.[9] If our null is rejected, then the specification of MPH models is rejected, so our test can used as a falsification test for MPH models.

5.2 Duration of strikes

In this subsection, we test the specification of GAFT models using data on the duration of strikes. Here Y is the duration of strikes in U.S. manufacturing firms, defined as the number of days since the start of a strike. Our X is a scalar variable indicator of the business cycle position of the economy, measured by the deviation of output from its trend. Positive values of X mean that the economy is above its growth trend. We assume that A.1 holds with $X \perp U$, i.e., Z is empty.

Our dataset was used in Kennan (1985) and is employed in several econometrics textbooks including as Cameron and Trivedi (2005) and Greene (2011). The sample size is 566. Table 5 presents data summary statistics.

[9] More broadly, this proposition shows that nonparametrically the only difference between GAFT and MPH models is some regularity conditions, since if one is given a GAFT model which by Ridder (1990) satisfies $Y = G[H_1(X) + U]$, then given the regularity assumed in Proposition 5.1, one can construct an equivalent MPH model by letting $\xi = [-\ln(1-\varepsilon)]\, e^{-U}$ where ε is uniform.

Table 5: Summary statistics for the strike data (sample size $n = 566$)

Variable	Name	Mean	Median	Standard Deviation	Min	Max
Y	Duration of strikes (days)	43.62	28.00	44.67	1	235
X	Business cycle position	0.006	0.008	0.050	-0.140	0.086

We apply our subsampling based test. The details of implementation is the same as in the simulations. For the bandwidth sequence b and c, we try various values of the constant κ, letting $\kappa = 0.5$, 0.75, 1, 1.25, 1.5 and 2. Results based on 1000 subsamples are reported in Table 6. Our results are robust, yielding similar p-values across different bandwidths and subsample sizes. The p-values are high for all subsample sizes under investigation. This suggests that our test supports the specification of GAFT models.

Table 6: p-values for the strike data (sample size $n = 566$)

Subsample size	$\lfloor n^{0.80} \rfloor = 159$	$\lfloor n^{0.85} \rfloor = 219$	$\lfloor n^{0.90} \rfloor = 300$
$\kappa = 0.5$	0.605	0.497	0.472
$\kappa = 0.75$	0.617	0.522	0.496
$\kappa = 1$	0.651	0.578	0.572
$\kappa = 1.25$	0.599	0.558	0.572
$\kappa = 1.5$	0.540	0.534	0.530
$\kappa = 2$	0.374	0.437	0.445

6 Extensions

Our methodology can be extended to test other related hypotheses for specifications in nonseparable models. For example, suppose that X is multi-dimensional such that $X \equiv (X_1, X_2)$. Then our results can be used to test the hypotheses:

$$\begin{aligned}
\mathbb{H}_{20} :\ & \text{There exist two measurable functions } R_2 \text{ and } H_3 \text{ such that} \\
& \quad Y = R_2 \left[H_3 \left(X_1, X_2 \right), U \right] \ a.s. \\
\mathbb{H}_{2A} :\ & \mathbb{H}_{20} \text{ is false;}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{H}_{30} :\ & \text{There exist three measurable functions } R_3,\ H_4 \text{ and } H_5 \text{ such that} \\
& \quad Y = R_3 \left[H_4 \left(X_1 \right) + H_5 \left(X_2 \right), U \right] \ a.s. \\
\mathbb{H}_{3A} :\ & \mathbb{H}_{30} \text{ is false.}
\end{aligned}$$

Given the key conditional exogeneity assumption A.1, a testable implication of $\mathbb{H}_{20}$ is

$$\frac{\partial F_{Y|X_1,X_2,Z} \left(y \mid x_1, x_2, z \right) / \partial x_1}{\partial F_{Y|X_1,X_2,Z} \left(y \mid x_1, x_2, z \right) / \partial x_2} = r_3 \left(x_1, x_2 \right), \tag{6.1}$$

where $F_{Y|X_1,X_2,Z} \left(y \mid x_1, x_2, z \right)$ is the conditional CDF of Y given (X_1, X_2, Z) and r_3 some unknown measurable function. Similarly, $\mathbb{H}_{30}$ implies that

$$\frac{\partial F_{Y|X_1,X_2,Z} \left(y \mid x_1, x_2, z \right) / \partial x_1}{\partial F_{Y|X_1,X_2,Z} \left(y \mid x_1, x_2, z \right) / \partial x_2} = r_4 \left(x_1 \right) \cdot r_5 \left(x_2 \right) \tag{6.2}$$

for some unknown measurable functions r_4 and r_5.

Our test can also be extended to test for semiparametric specifications. For example, one may be interested in testing

$$
\begin{aligned}
\mathbb{H}_{40}: & \text{ There exist } \beta \in \mathbb{R}^{d_x} \text{ and two measurable functions } R_4 \text{ and } H_2 \text{ such that} \\
& \quad Y = R_4\left[X'\beta + H_2(U)\right] \ a.s., \text{ and } R_4 \text{ is strictly monotonic.} \\
\mathbb{H}_{4A}: & \ \mathbb{H}_{40} \text{ is false;}
\end{aligned}
$$

Then $\mathbb{H}_{40}$ implies that

$$\frac{D_x F(y \mid x, z)}{f(y \mid x, z)} = r_6(y) \tag{6.3}$$

for some unknown measurable function $r_{6.}$

To test equations (6.1), (6.2), and (6.3), one can readily construct test statistics similar to ours, using marginal integration as proposed in testing $\mathbb{H}_0$.

7 Concluding Remarks

In this paper, we proposed a specification test for a transformation model containing a vector of covariates and a vector of unobservable errors. This test is related to tests for separability and monotonicity in nonseparable structural equations. We exploit the testable implication of the transformation model that the ratio of the derivatives of a conditional CDF takes a product form. Our test statistics are based on the L_2 distance between restricted and unrestricted estimators of this ratio of derivatives. We show that the test statistics are asymptotically normal and consistent against the alternative of this testable implication. We provide limit normal distribution theory as well as subsampling methods for obtaining p-values under the null. Our simulations suggest that the test statistics perform well in moderate size samples. We apply our statistic to test the specification of generalized accelerated failure-time (GAFT) models for data on the durations of strikes among manufacturers in the US and fail to reject the specification of GAFT models. We find this result to be stable and robust over a wide range of tuning parameter values

Appendix

A Proof of the main results in Section 2

Proof of Theorem 2.1. We first prove (a). Let $\tilde{U} = H_2(U)$. Then

$$\begin{aligned}
F(y \mid x, z) &= \Pr[Y \leq y \mid X = x, Z = z] \\
&= \Pr\left[G\left[H_1(X) + \tilde{U}\right] \leq y \mid X = x, Z = z\right] \\
&= \Pr\left[\tilde{U} \leq G^{-1}(y) - H_1(x) \mid Z = z\right] \\
&= F_{\tilde{U}|Z}\left[G^{-1}(y) - H_1(x), z\right],
\end{aligned}$$

where $F_{\tilde{U}|Z}(\cdot, z)$ denotes the conditional CDF of $\tilde{U}$ given $Z = z$. Let $F_{1,\tilde{U}|Z}$ be the derivative of $F_{\tilde{U}|Z}$ with respect to its first argument. Then,

$$\frac{\partial F(y \mid x, z)/\partial x}{\partial F(y \mid x, z)/\partial y} = \frac{F_{1,\tilde{U}|Z}\left[G^{-1}(y) - H_1(x), z\right] \cdot [-\partial H_1(x)/\partial x]}{F_{1,\tilde{U}|Z}\left[G^{-1}(y) - H_1(x), z\right] \cdot [\partial G^{-1}(y)/\partial y]} = \frac{-\partial H_1(x)/\partial x}{\partial G^{-1}(y)/\partial y}.$$

So the functions s_1 and s_2 exist and are given by $s_1(x) = -C\, \partial H_1(x)/\partial x$ and $s_2(y) = \frac{1}{C \partial G^{-1}(y)/\partial y}$, where $C \neq 0$ is an arbitrary constant. Clearly, $s_2 : \mathbb{R} \to \mathbb{R}_+$ if $C > 0$ and $s_2 : \mathbb{R} \to \mathbb{R}_-$ if $C < 0$. The measurable functions S_1 and S_2 are given by CH_1 and CG^{-1}, respectively.

We now prove (b). Without loss of generality, assume that $s_2 : \mathbb{R} \to \mathbb{R}_+$. We can always find two scalar functions S_1 and S_2 such that $\partial S_1(x)/\partial x = -s_1(x)$ and $\partial S_2(y)/\partial y = 1/s_2(y)$, where $S_2(\cdot)$ is strictly increasing. Combining this with the definition of $r(y; x, z)$ gives

$$\frac{D_x F(y \mid x, z)}{D_y F(y \mid x, z)} = s_1(x)\, s_2(y) = \frac{-\partial S_1(x)/\partial x}{\partial S_2(y)/\partial y} \text{ for all } (x, y, z) \in \mathcal{W}. \tag{A.1}$$

Let $\tilde{U} \equiv S_2(Y) - S_1(X)$ and $\tilde{u} \equiv S_2(y) - S_1(x)$. By the monotonicity of S_2, we have $Y = S_2^{-1}[S_1(X) + \tilde{U}]$ and $y = S_2^{-1}[S_1(x) + \tilde{u}]$. It follows that

$$\begin{aligned}
F_{\tilde{U}|X,Z}(\tilde{u}, x, z) &\equiv P\left(\tilde{U} \leq \tilde{u} | X = x, Z = z\right) = P\left(S_2(Y) - S_1(X) \leq \tilde{u} | X = x, Z = z\right) \\
&= P\left(Y \leq S_2^{-1}\left(S_1(x) + \tilde{u}\right) | X = x, Z = z\right) \\
&= P\left(Y \leq y | X = x, Z = z\right) = F(y|x, z).
\end{aligned}$$

Then

$$\begin{aligned}
\frac{D_x F(y \mid x, z)}{D_y F(y \mid x, z)} &= \frac{D_x F_{\tilde{U}|X,Z}(\tilde{u}, x, z)}{D_y F_{\tilde{U}|X,Z}(\tilde{u}, x, z)} = \frac{\partial F_{\tilde{U}|X,Z}(\tilde{u}, x, z)/\partial \tilde{u} \cdot s_1(x) + \partial F_{\tilde{U}|X,Z}(\tilde{u}, x, z)/\partial x}{\partial F_{\tilde{U}|X,Z}(\tilde{u}, x, z)/\partial \tilde{u} \cdot (1/s_2(y))} \\
&= s_1(x)\, s_2(y) + \frac{\partial F_{\tilde{U}|X,Z}(\tilde{u}, x, z)/\partial x \cdot (1/s_2(y))}{\partial F_{\tilde{U}|X,Z}(\tilde{u}, x, z)/\partial \tilde{u}} \text{ for all } (x, y, z) \in \mathcal{W}.
\end{aligned} \tag{A.2}$$

Comparing (A.1) with (A.2) yields $\partial F_{\tilde{U}|X,Z}(\tilde{u}, x, z)/\partial x = 0$ for all $(\tilde{u}, x, z) \in \mathcal{U} \times \mathcal{V}$ where $\mathcal{U}$ denotes the support of $\tilde{U}$. Therefore, $\tilde{U} \perp X|Z$. So far, we have shown that

$$Y = S_2^{-1}[S_1(X) + \tilde{U}]$$

where S_2^{-1} is strictly monotonic and $X \perp \tilde{U}|Z$. The conclusion in part (b) follows by setting $G = S_2^{-1}$ and $H_1 = S_1$. ■

Proof of Corollary 2.2. Under $\mathbb{H}_{10}$ and Assumption A.1, (2.1) in Theorem 2.1(a) holds, implying that

$$\begin{aligned} r_0 &\equiv E_Y E_{XZ}\left[r\left(Y;X,Z\right)\right] = E\left[s_1\left(X\right)\right]E\left[s_2\left(Y\right)\mathbf{1}\left\{Y\in\mathcal{Y}_0\right\}\right], \\ r_1\left(x\right) &\equiv E\left[r\left(Y;x,Z\right)\right] = s_1\left(x\right)E\left[s_2\left(Y\right)\mathbf{1}\left\{Y\in\mathcal{Y}_0\right\}\right], \\ r_2\left(y\right) &\equiv E\left[r\left(y;X,Z\right)\right] = E\left[s_1\left(X\right)\right]s_2\left(y\right)\mathbf{1}\left\{y\in\mathcal{Y}_0\right\}. \end{aligned}$$

It follows that $r\left(Y;X,Z\right)\circ r_0-r_1\left(X\right)\circ r_2\left(Y\right)=\left[s_1\left(X\right)s_2\left(Y\right)\mathbf{1}\left\{Y\in\mathcal{Y}_0\right\}\right]\circ\left\{E\left[s_1\left(X\right)\right]E\left[s_2\left(Y\right)\mathbf{1}\left\{Y\in\mathcal{Y}_0\right\}\right]\right\}$ $-\left\{s_1\left(X\right)E\left[s_2\left(Y\right)\mathbf{1}\left\{Y\in\mathcal{Y}_0\right\}\right]\right\}\circ\left\{E\left[s_1\left(X\right)\right]s_2\left(Y\right)\mathbf{1}\left\{y\in\mathcal{Y}_0\right\}\right\}=0.$ ■

B Proof of the main results in Section 3

To prove Theorem 3.1, we first establish some technical lemmas. Recall that $V_i\equiv\left(X_i',Z_i'\right)'$, $v\equiv\left(x',z'\right)'$, $K_b\left(v\right)\equiv b^{-d}K\left(v/b\right)$, and $\mu_b\left(v\right)\equiv\mu\left(v/b\right)$. Let $W_i\equiv\left(Y_i,V_i'\right)'$ and $w\equiv\left(y,v'\right)'$. Define

$$\mathbf{B}_b\left(y;v\right)\equiv\frac{1}{n}\sum_{i=1}^{n}K_b\left(V_i-v\right)\mu_b\left(V_i-v\right)\Delta_{i,y}\left(v\right)\ \text{ and }\ \mathbf{V}_b\left(y;v\right)\equiv\frac{1}{n}\sum_{i=1}^{n}K_b\left(V_i-v\right)\mu_b\left(V_i-v\right)\bar{\mathbf{1}}_y\left(W_i\right),$$

where $\Delta_{i,y}\left(v\right)\equiv F\left(y|V_i\right)-F\left(y|v\right)-\sum_{1\le|\mathbf{j}|\le p}\frac{1}{\mathbf{j}!}D^{\mathbf{j}}F\left(y|v\right)\left(V_i-v\right)^{\mathbf{j}}$, and $\bar{\mathbf{1}}_y\left(W_i\right)\equiv\mathbf{1}\left\{Y_i\le y\right\}-F\left(y|V_i\right)$. Let $\bar{\mathbf{S}}_b\left(v\right)\equiv E[\mathbf{S}_b\left(v\right)]$ and $\bar{\mathbf{B}}_b\left(y;v\right)\equiv E[\mathbf{B}_b\left(y;v\right)]$, where $\mathbf{S}_b\left(v\right)$ is defined after (2.13). The next lemma establishes uniform consistency of $D_x\hat{F}_b\left(y|v\right)$.

Lemma B.1 *Suppose that Assumptions C.1-C.3, C.5, and C.7 hold. Then uniformly in* $(y,v)\in\mathcal{Y}_0\times\mathcal{V}$,
(a) $D_x\hat{F}_b\left(y|v\right)-D_xF\left(y|v\right)=b^{-1}e_1\bar{\mathbf{S}}_b\left(v\right)^{-1}\left[\mathbf{V}_b\left(y;v\right)+\bar{\mathbf{B}}_b\left(y;v\right)\right]+O_P(\nu_{1b}^2b^{-1}+\nu_{1b}b^p)$,
(b) $D_x\hat{F}_b\left(y|v\right)-D_xF\left(y|v\right)=O_P(\nu_{1b}b^{-1}+b^p)$,
where $\nu_{1b}\equiv n^{-1/2}b^{-d/2}\sqrt{\ln n}$.

Proof. By Lemma 10.1 in HSW (2011), $\hat{\boldsymbol{\beta}}\left(y|v\right)-\boldsymbol{\beta}\left(y|v\right)=\bar{\mathbf{S}}_b\left(v\right)^{-1}\left[\mathbf{V}_b\left(y;v\right)+\bar{\mathbf{B}}_b\left(y;v\right)\right]+O_P(\nu_{1b}^2+\nu_{1b}b^{p+1})=O_P(\nu_{1b}+b^{p+1})$ uniformly in $(y,v)\in\mathcal{Y}_0\times\mathcal{V}$. The results follow from the fact hat $D_x\hat{F}_b\left(y|v\right)-D_xF\left(y|v\right)=e_1[\hat{\boldsymbol{\beta}}\left(y|x,z\right)-\boldsymbol{\beta}\left(y|v\right)]/b$. ■

Define

$$\begin{aligned} \mathbf{V}_c^{(L)}\left(y;v\right) &\equiv \frac{1}{n}\sum_{i=1}^{n}K_c\left(V_i-v\right)\mu_c\left(V_i-v\right)\bar{\mathbf{L}}_y\left(W_i\right), \\ \mathbf{B}_c^{(L)}\left(y;v\right) &\equiv \frac{1}{n}\sum_{i=1}^{n}K_c\left(V_i-v\right)\mu_c\left(V_i-v\right)\left[\alpha\left(y|V_i\right)-f\left(y|v\right)-\sum_{1\le|\mathbf{j}|\le p}\frac{1}{\mathbf{j}!}\alpha^{(\mathbf{j})}\left(y|v\right)\left(V_i-v\right)^{\mathbf{j}}\right], \end{aligned}$$

where $\bar{\mathbf{L}}_y\left(W_i\right)\equiv L_c(Y_i-y)-\alpha\left(y|V_i\right)$ and $\alpha\left(y|v\right)\equiv E\left[L_c(Y_i-y)|V_i=v\right]$. Let $\bar{\mathbf{B}}_c^{(L)}\left(y;v\right)\equiv E[\mathbf{B}_c^{(L)}\left(y;v\right)]$. The next lemma establishes uniform consistency of $\hat{f}_c(y|v)$.

Lemma B.2 *Suppose that Assumptions C.1-C.7 hold. Then uniformly in* $(y,v)\in\mathcal{Y}_0\times\mathcal{V}$,
(a) $\hat{f}_c(y|v)-f(y|v)=e_2'\bar{\mathbf{S}}_c\left(v\right)^{-1}\left[\mathbf{V}_c^{(L)}\left(y;v\right)+\bar{\mathbf{B}}_c^{(L)}\left(y;v\right)\right]+O_P(\nu_{2c}^2+\nu_{2c}c^{p+1})$,
(b) $\hat{f}_c(y|v)-f(y|v)=O_P(\nu_{2c}+c^{p+1}+c^r)$,
where $\nu_{2c}\equiv n^{-1/2}c^{-(d+1)/2}\sqrt{\ln n}$.

Proof. The results follow from Lemma 10.5 in HSW (2011) who prove the results based on standard arguments as used in Masry (1996), Hansen (2008), and Kong et al. (2010). ■

Lemma B.3 *Suppose that Assumptions C.1-C.7 hold. Then*

(a) $\hat{r}(y;v) - r(y;v) = b^{-1} e_1 \bar{\mathbf{S}}_b(v)^{-1} \mathbf{V}_b(y;v) f(y|v)^{-1} - D_x F(y|v) e_2' \bar{\mathbf{S}}_c(v)^{-1} \mathbf{V}_c^{(L)}(y;v) f(y|v)^{-2} + O_P(\nu_{bc})$ *uniformly in* $(y,v) \in \mathcal{Y}_0 \times \mathcal{V}$,

(b) $\hat{r}_0 - r_0 = O_P\left(\nu_{bc} + n^{-1/2} b^{-1}\right)$,

(c) $\sup_{y \in \mathcal{Y}_0} |\hat{r}_2(y) - r_2(y)| = O_P(\nu_{bc} + n^{-1/2} b^{-1} \sqrt{\ln n})$,

where $\nu_{bc} \equiv \nu_{1b}^2 b^{-1} + b^p + \nu_{2c}^2 + c^{p+1} + c^r + \nu_{1b}\nu_{2c} b^{-1}$.

Proof. (a) Let $\hat{q}(y;v) \equiv \hat{r}(y;v) - r(y;v)$. Noting that $\hat{f}_c(y|v)^{-1} = f(y|v)^{-1} - [\hat{f}_c(y|v) - f(y|v)]/f(y|v)^2 + R_1(y;v)$ where $R_1(y;v) \equiv [\hat{f}_c(y|v) - f(y|v)]^2/[f(y|v)^2 \hat{f}_c(y|v)]$, we have that for any $(y,v) \in \mathcal{Y}_0 \times \mathcal{V}$,

$$\begin{aligned}
\hat{q}(y;v) &= \frac{D_x \hat{F}_b(y|v)}{\hat{f}_c(y|v)} - \frac{D_x F(y|v)}{f(y|v)} = \frac{D_x \hat{F}_b(y|v) - D_x F(y|v)}{f(y|v)} + \left[\frac{1}{\hat{f}_c(y|v)} - \frac{1}{f(y|v)}\right] D_x F(y|v) + R_2(y;v) \\
&= \frac{D_x \hat{F}_b(y|v) - D_x F(y|v)}{f(y|v)} - \frac{\hat{f}_c(y|v) - f(y|v)}{f(y|v)^2} D_x F(y|v) + R_1(y;v) D_x F(y|v) + R_2(y;v) \\
&\equiv \hat{q}_1(y;v) - \hat{q}_2(y;v) + R_1(y;v) D_x F(y|v) + R_2(y;v), \text{ say,}
\end{aligned}$$

where $R_2(y;v) \equiv [\hat{f}_c(y|v)^{-1} - f(y|v)^{-1}]\left[D_x \hat{F}_b(y|v) - D_x F(y|v)\right]$. Using Lemmas B.1 and B.2 we can bound the last two terms in the last expression uniformly by $O_p(\eta_{2c}(\eta_{1b} + \eta_{2c}))$, where $\eta_{1b} \equiv \nu_{1b} b^{-1} + b^p$ and $\eta_{2c} \equiv \nu_{2c} + c^{p+1} + c^r$. In addition, uniformly in $(y,v) \in \mathcal{Y}_0 \times \mathcal{V}$,

$$\begin{aligned}
\hat{q}_1(y;v) &= [D_x \hat{F}_b(y|v) - D_x F(y|v)]/f(y|v) \\
&= b^{-1} e_1 \bar{\mathbf{S}}_b(v)^{-1} [\mathbf{V}_b(y;v) + \bar{\mathbf{B}}_b(y;v)]/f(y|v) + O_P(\nu_{1b}^2 b^{-1} + \nu_{1b} b^p) \\
&= b^{-1} e_1 \bar{\mathbf{S}}_b(v)^{-1} \mathbf{V}_b(y;v)/f(y|v) + O_P(\nu_{1b}^2 b^{-1} + b^p),
\end{aligned}$$

and

$$\begin{aligned}
\hat{q}_2(y;v) &= D_x F(y|v) [\hat{f}_c(y|v) - f(y|v)]/f(y|v)^2 \\
&= D_x F(y|v) e_2' \bar{\mathbf{S}}_c(v)^{-1} [\mathbf{V}_c^{(L)}(y;v) + \bar{\mathbf{B}}_c^{(L)}(y;v)]/f(y|v)^2 + O_P(\nu_{2c}^2 + \nu_{2c} c^{p+1}) \\
&= D_x F(y|v) e_2' \bar{\mathbf{S}}_c(v)^{-1} \mathbf{V}_c^{(L)}(y;v)/f(y|v)^2 + O_P(\nu_{2c}^2 + c^{p+1} + c^r).
\end{aligned}$$

It follows that uniformly in $(y,v) \in \mathcal{Y}_0 \times \mathcal{V}$,

$$\begin{aligned}
\hat{q}(y;v) &= b^{-1} e_1 \bar{\mathbf{S}}_b(v)^{-1} \mathbf{V}_b(y;v) f(y|v)^{-1} - D_x F(y|v) e_2' \bar{\mathbf{S}}_c(v)^{-1} \mathbf{V}_c^{(L)}(y;v) f(y|v)^{-2} \\
&\quad + O_P\left(\nu_{1b}^2 b^{-1} + b^p + \nu_{2c}^2 + c^{p+1} + c^r + \eta_{2c}(\eta_{1b} + \eta_{2c})\right) \\
&= b^{-1} e_1 \bar{\mathbf{S}}_b(v)^{-1} \mathbf{V}_b(y;v) f(y|v)^{-1} - D_x F(y|v) e_2' \bar{\mathbf{S}}_c(v)^{-1} \mathbf{V}_c^{(L)}(y;v) f(y|v)^{-2} + O_P(\nu_{bc}).
\end{aligned}$$

(b) Write $\hat{r}_0 - r_0 = \hat{r}_{01} + \hat{r}_{02}$, where $\hat{r}_{01} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [\hat{r}(Y_i; X_j, Z_j) - r(Y_i; X_j, Z_j)]$ and $\hat{r}_{02} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [r(Y_i; X_j, Z_j) - r_0]$. It is easy to show that $\hat{r}_{02} = O_P\left(n^{-1/2}\right)$ by the Chebyshev inequality. For $\hat{r}_{01}$, we have by (a) that $\hat{r}_{01} = R_{1n} - R_{2n} + O_P(\nu_{bc})$, where $R_{1n} \equiv \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mathbf{V}_b(Y_i; V_j) \times f_{ij}^{-1} \mathbf{1}_i$, $R_{2n} \equiv \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n D_{xij} e_2' \bar{\mathbf{S}}_c(V_j)^{-1} \mathbf{V}_c^{(L)}(Y_i; V_j) f_{ij}^{-2} \mathbf{1}_i$, $\mathbf{1}_i \equiv \mathbf{1}\{Y_i \in \mathcal{Y}_0\}$, $f_{ij} \equiv f(Y_i|V_j)$, $f_i \equiv f(Y_i|V_i)$ and $D_{xij} \equiv D_x F(Y_i|V_j)$. For R_{1n}, we have

$$\begin{aligned}
R_{1n} &= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mu_b(V_k - V_j) K_b(V_k - V_j) \bar{\mathbf{1}}_{Y_i}(W_k) f_{ij}^{-1} \mathbf{1}_i \\
&= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \sum_{k=1, k \neq j, i}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mu_b(V_k - V_j) K_b(V_k - V_j) \bar{\mathbf{1}}_{Y_i}(W_k) f_{ij}^{-1} \mathbf{1}_i \\
&\quad + \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1, j \neq i}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mu_b(V_i - V_j) K_b(V_i - V_j) \bar{\mathbf{1}}_{Y_i}(W_i) f_{ij}^{-1} \mathbf{1}_i
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} b^{-1}e_1\bar{\mathbf{S}}_b(V_j)^{-1}\mu_b(0)K_b(0)\bar{\mathbf{1}}_{Y_i}(W_j)f_{ij}^{-1}\mathbf{1}_i \\
&+\frac{1}{n^3}\sum_{i=1}^{n}\sum_{k=1,i}^{n} b^{-1}e_1\bar{\mathbf{S}}_b(V_i)^{-1}\mu_b(V_k-V_i)K_b(V_k-V_i)\bar{\mathbf{1}}_{Y_i}(W_k)f_i^{-1}\mathbf{1}_i \\
&+\frac{1}{n^3}\sum_{i=1}^{n} b^{-1}e_1\bar{\mathbf{S}}_b(V_i)^{-1}\mu_b(0)K_b(0)\bar{\mathbf{1}}_{Y_i}(W_i)f_i^{-1}\mathbf{1}_i \\
\equiv\ & R_{1n,1}+R_{1n,2}+R_{1n,3}+R_{1n,4}+R_{1n,5}.
\end{aligned}
$$

It is easy to show that $R_{1n,5}=O_P\left(n^{-2}b^{-d-1}\right)$, $R_{1n,4}=O_P\left(n^{-3/2}b^{-1}\right)$, $R_{1n,3}=O_P\left(n^{-3/2}b^{-d-1}\right)$, and $R_{1n,2}=O_P\left(n^{-1}b^{-1}\right)$. Noting that $R_{1n,1}$ is a third-order U-statistic with $E\left(R_{n,1}\right)=0$, it is straightforward to show that $E\left(R_{1n,1}^2\right)=O\left(n^{-1}b^{-2}+n^{-2}b^{-d-2}\right)$. Thus $R_{1n,1}=O_P\left(n^{-1/2}b^{-1}\right)$ and $R_{1n}=O_P\left(n^{-1/2}b^{-1}\right)$ as $n^{-1}b^{-d}=o(1)$. By the same token, we can show that $R_{2n}=O_P\left(n^{-1/2}\right)$. It follows that $\hat{r}_0-r_0=O_P(\nu_{bc}+n^{-1/2}b^{-1})$.

(c) Write $\hat{r}_2(y)-r_2(y)=\hat{r}_{21}(y)+\hat{r}_{22}(y)$, where $\hat{r}_{21}(y)=\frac{1}{n}\sum_{i=1}^{n}\left[\hat{r}(y;V_i)-r(y;V_i)\right]$ and $\hat{r}_{22}(y)=\frac{1}{n}\sum_{i=1}^{n}\left[r(y;V_i)-r_2(y)\right]$. By standard chaining arguments and the exponential inequality, we can show that $\sup_{y\in\mathcal{Y}_0}\|\hat{r}_{22}(y)\|=O(n^{-1/2}\sqrt{\ln n})$. By (a), $\hat{r}_{21}(y)=\bar{r}_{21}(y)+O_P(\nu_{bc})$ uniformly in $y\in\mathcal{Y}_0$, where $\bar{r}_{21}(y)\equiv\bar{r}_{21,1}(y)-\bar{r}_{21,2}(y)$, $\bar{r}_{21,1}(y)\equiv\frac{1}{n}\sum_{i=1}^{n}b^{-1}e_1\bar{\mathbf{S}}_b(V_i)^{-1}\mathbf{V}_b(y;V_i)f(y|V_i)^{-1}\mathbf{1}\{y\in\mathcal{Y}_0\}$, and $\bar{r}_{21,2}(y)\equiv\frac{1}{n}\sum_{i=1}^{n}f(y|V_i)^{-2}D_xF(y|V_i)\times e_2'\bar{\mathbf{S}}_c(V_i)^{-1}\mathbf{V}_c^{(L)}(y;V_i)\mathbf{1}\{y\in\mathcal{Y}_0\}$. Now write $\bar{r}_{21,1}(y)$ as the summation of a first order U-statistic and a second order U-statistic: $\bar{r}_{21,1}(y)=\bar{r}_{21,11}(y)+\bar{r}_{21,12}(y)$, where

$$
\begin{aligned}
\bar{r}_{21,11}(y) &\equiv \frac{1}{n^2}\sum_{i=1}^{n}b^{-1}e_1\bar{\mathbf{S}}_b(V_i)^{-1}\mu_b(0)K_b(0)\bar{\mathbf{1}}_y(W_i)f(y|V_i)^{-1}\mathbf{1}\{y\in\mathcal{Y}_0\},\ \text{and} \\
\bar{r}_{21,12}(y) &\equiv \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}b^{-1}e_1\bar{\mathbf{S}}_b(V_i)^{-1}\mu_b(V_j-V_i)K_b(V_j-V_i)\bar{\mathbf{1}}_y(W_j)f(y|V_i)^{-1}\mathbf{1}\{y\in\mathcal{Y}_0\}.
\end{aligned}
$$

By the exponential inequality, we can show that $\sup_{y\in\mathcal{Y}_0}\|\bar{r}_{21,11}(y)\|=O(n^{-3/2}b^{-d-1}\sqrt{\ln n})$. For $\bar{r}_{21,12}(y)$, one can follow the proof of (A.10) in Gozalo and Linton (2001) and show that $\sup_{y\in\mathcal{Y}_0}||\bar{r}_{21,12}(y)||=O(n^{-1/2}b^{-1}\sqrt{\ln n})$.[10] Hence $\sup_{y\in\mathcal{Y}_0}\|\bar{r}_{21,1}(y)\|=O(n^{-1/2}b^{-1}\sqrt{\ln n})$. Similarly, $\sup_{y\in\mathcal{Y}_0}\|\bar{r}_{21,2}(y)\|=O(n^{-1/2}\sqrt{\ln n})$. Thus $\sup_{y\in\mathcal{Y}_0}\|\hat{r}_2(y)-r_2(y)\|=O_P(\nu_{bc}+n^{-1/2}b^{-1}\sqrt{\ln n})$. ■

Proof of Theorem 3.1. Let $a_i\equiv a(Y_i;X_i,Z_i)$, $r_i\equiv r(Y_i;X_i,Z_i)$, $r_{1i}\equiv r_1(X_i)$, $r_{2i}\equiv r_2(Y_i)$, $\hat{r}_i\equiv\hat{r}(Y_i;X_i,Z_i)$, $\hat{r}_{1i}\equiv\hat{r}_1(X_i)$, and $\hat{r}_{2i}\equiv\hat{r}_2(Y_i)$. Let $\xi_{1i}=r_i\circ r_0-r_{1i}\circ r_{2i}$, $\xi_{2i}=(\hat{r}_i-r_i)\circ r_0+r_i\circ(\hat{r}_0-r_0)-(\hat{r}_{1i}-r_{1i})\circ r_{2i}-r_{1i}\circ(\hat{r}_{2i}-r_{2i})$, and $\xi_{3i}=(\hat{r}_i-r_i)\circ(\hat{r}_0-r_0)-(\hat{r}_{1i}-r_{1i})\circ(\hat{r}_{2i}-r_{2i})$. Then

$$
\begin{aligned}
nb^{\frac{d}{2}+2}\hat{\Gamma} &= b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left\|\left[(\hat{r}_i-r_i)+r_i\right]\circ\left[(\hat{r}_0-r_0)+r_0\right]-\left[(\hat{r}_{1i}-r_{1i})+r_{1i}\right]\circ\left[(\hat{r}_{2i}-r_{2i})+r_{2i}\right]\right\|^2 a_i \\
&= b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left\|\xi_{1i}+\xi_{2i}+\xi_{3i}\right\|^2 a_i \\
&= \Gamma_{1n}+\Gamma_{2n}+\Gamma_{3n}+2\Gamma_{4n}+2\Gamma_{5n}+2\Gamma_{6n}, \qquad \text{(B.1)}
\end{aligned}
$$

[10] If we ignore the boundary points, we can write $\bar{\mathbf{S}}_b(v)=f(v)\mathbb{S}+b\mathbb{V}(v)+o(b)$ uniformly in v in the interior of $\mathcal{V}$, where $\mathbb{S}$ and $\mathbb{V}$ are defined as M and V in Li et al. (2003, p. 617). Following the proof of Lemma A.3 in their paper, one can show that $\bar{r}_{21,12}(y)=O(n^{-1/2})$ elementwise by using the degeneracy of the second order U-statistic defined analogously to $\bar{r}_{21,12}(y)$ but with $\bar{\mathbf{S}}_b(V_i)^{-1}$ replaced by its leading term $f(V_i)^{-1}\mathbb{S}^{-1}$. But their argument breaks down when v takes values on the boundary of $\mathcal{V}$.

where $\Gamma_{ln} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\|\xi_{li}\|^2 a_i$ for $l=1$, 2, and 3, $\Gamma_{4n} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\xi_{1i}'\xi_{2i}a_i$, $\Gamma_{5n} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\xi_{1i}'\xi_{3i}a_i$, and $\Gamma_{6n} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\xi_{2i}'\xi_{3i}a_i$. Under $\mathbb{H}_0$, $\Gamma_{jn}=0$ for $j=1$, 4, 5. It suffices to prove the theorem by showing that (i) $\Gamma_{2n}-\mathbb{B}_n \xrightarrow{d} N\left(0,\sigma_0^2\right)$, (ii) $\Gamma_{3n}=o_P(1)$, and (iii) $\Gamma_{6n}=o_P(1)$.

To show (i), we write $\Gamma_{2n}=\sum_{j=1}^{10}\Gamma_{2n,j}$ where $\Gamma_{2n,1}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\|(\hat{r}_i-r_i)\circ r_0\|^2 a_i$, $\Gamma_{2n,2}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\|r_i\circ(\hat{r}_0-r_0)\|^2 a_i$, $\Gamma_{2n,3}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\|(\hat{r}_{1i}-r_{1i})\circ r_{2i}\|^2 a_i$, $\Gamma_{2n,4}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\|r_{1i}\circ(\hat{r}_{2i}-r_{2i})\|^2 a_i$, $\Gamma_{2n,5}\equiv 2b^{\frac{d}{2}+2}\sum_{i=1}^{n}((\hat{r}_i-r_i)\circ r_0)'(r_i\circ(\hat{r}_0-r_0))\,a_i$, $\Gamma_{2n,6}\equiv -2b^{\frac{d}{2}+2}\sum_{i=1}^{n}((\hat{r}_i-r_i)\circ r_0)'((\hat{r}_{1i}-r_{1i})\circ r_{2i})\,a_i$, $\Gamma_{2n,7}\equiv -2b^{\frac{d}{2}+2}\sum_{i=1}^{n}((\hat{r}_i-r_i)\circ r_0)'(r_{1i}\circ(\hat{r}_{2i}-r_{2i}))\,a_i$, $\Gamma_{2n,8}\equiv -2b^{\frac{d}{2}+2}\sum_{i=1}^{n}(r_i\circ(\hat{r}_0-r_0))'((\hat{r}_{1i}-r_{1i})\circ r_{2i})a_i$, $\Gamma_{2n,9}\equiv -2b^{\frac{d}{2}+2}\sum_{i=1}^{n}(r_i\circ(\hat{r}_0-r_0))'(r_{1i}\circ(\hat{r}_{2i}-r_{2i}))\,a_i$, and $\Gamma_{2n,10}\equiv 2b^{\frac{d}{2}+2}\sum_{i=1}^{n}((\hat{r}_{1i}-r_{1i})\circ r_{2i})'(r_{1i}\circ(\hat{r}_{2i}-r_{2i}))a_i$. We show in Lemma B.4 below that $\Gamma_{2n,1}$ contributes to both the asymptotic bias and variance of our test statistic: $\Gamma_{2n,1}-\mathbb{B}_{1n}\xrightarrow{d} N\left(0,\sigma_0^2\right)$, and $\mathbb{B}_{1n}=O_P(b^{\frac{d}{2}+2}\left(b^{-d-2}+c^{-d-1}\right))$ is the dominant bias term that never vanishes asymptotically no matter whether Z is absent or not. The normalization constant $nb^{\frac{d}{2}+2}$ in the front of $\hat{\Gamma}$ in (B.1) indicates that the rate of $\hat{\Gamma}$ converging to zero under the null is uniquely determined by the convergence rate of $\hat{r}_i$ to r_i as the asymptotic variance of $\hat{r}(y;x,z)-r(y;x,z)$ is of the order $O\left(\left(n^{-1}b^{-(d+2)}\right)\right)$. We show in Lemmas B.5(b) and B.6(c) below that $\Gamma_{2n,3}$ and $\Gamma_{2n,6}$ also contribute to the asymptotic bias of our test statistic: $\Gamma_{2n,3}=\mathbb{B}_{2n}+o_P(1)$ and $\Gamma_{2n,6}=-2\mathbb{B}_{3n}+o_P(1)$, $\mathbb{B}_{2n}=O_P(b^{(d_z-d_x)/2}+b^{\frac{d}{2}+2}c^{-d_x})$ and $\mathbb{B}_{3n}=O_P(b^{(d_z-d_x)/2}+b^{\frac{d}{2}+2}c^{-d_x})$. $\mathbb{B}_{2n}$ signifies the estimation effect of $\hat{r}_{1i}$ and $\mathbb{B}_{3n}$ signifies the estimation effect of both $\hat{r}_i$ and $\hat{r}_{1i}$. $\mathbb{B}_{2n}$ and $\mathbb{B}_{3n}$ become $o_P(1)$ if $d_z>d_x$ and $b^{\frac{d}{2}+2}c^{-d_x}=o(1)$. We also show in the other parts of Lemmas B.5(b) and B.6(c) below that $\Gamma_{2n,s}=o_P(1)$ for $s=2,4,5,7,...,10$.

In short, by Lemmas B.4, B.5(b) and B.6(c) below, $\Gamma_{2n,1}+\Gamma_{2n,3}+\Gamma_{2n,6}-\mathbb{B}_n\xrightarrow{d} N\left(0,\sigma_0^2\right)$ where $\mathbb{B}_n\equiv\mathbb{B}_{1n}+\mathbb{B}_{2n}-2\mathbb{B}_{3n}$, and by Lemmas B.5(a) and (c) and Lemmas B.6 (a) and (c)-(f), $\Gamma_{2n,s}=o_P(1)$ for $s=2,4,5,7,...,10$. It follows that $\Gamma_{2n}-\mathbb{B}_n\xrightarrow{d} N\left(0,\sigma_0^2\right)$.

Next, we show (ii). By Cauchy-Schwarz inequality and the fact that $\|A\circ B\|\leq\|A\|\,\|B\|$ for any two conformable vectors A and B, $\Gamma_{3n}\leq 2\Gamma_{3n,1}+2\Gamma_{3n,2}$, where $\Gamma_{3n,1}\equiv b^{\frac{d}{2}+2}\|\hat{r}_0-r_0\|^2\sum_{i=1}^{n}\|\hat{r}_i-r_i\|^2 a_i$ and $\Gamma_{3n,2}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\|(\hat{r}_{1i}-r_{1i})\circ(\hat{r}_{2i}-r_{2i})\|^2 a_i$. Following the arguments used in the proofs of Lemma B.4 and Lemma B.5(b) respectively, we can readily show that

$$\bar{\Gamma}_{3n,1} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\|\hat{r}_i-r_i\|^2 a_i = O_P\left(b^{\frac{d}{2}+2}\left(b^{-d-2}+c^{-d-1}\right)\right) \text{ and} \tag{B.2}$$

$$\bar{\Gamma}_{3n,2} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\|\hat{r}_{1i}-r_{1i}\|^2 a_i = O_P(b^{\frac{d}{2}+2}(b^{-d_x-2}+c^{-d_x})). \tag{B.3}$$

Then by Lemma B.3(b), $\Gamma_{3n,1}=\|\hat{r}_0-r_0\|^2\bar{\Gamma}_{3n,1}=O_P\left(\nu_{bc}^2+n^{-1}b^{-2}\right)O_P(b^{\frac{d}{2}+2}\left(b^{-d-2}+c^{-d-1}\right))=o_P(1)$ and $\Gamma_{3n,2}\leq\sup_{y\in\mathcal{Y}_0}\|\hat{r}_2(y)-r_2(y)\|^2\bar{\Gamma}_{3n,2}=O_P\left(\nu_{bc}^2+n^{-1}b^{-2}\ln n\right)O_P(b^{\frac{d}{2}+2}\left(b^{-d_x-2}+c^{-d_x}\right))=o_P(1)$. Consequently, $\Gamma_{3n}=o_P(1)$.

To show (iii), note that $\Gamma_{6n}=\Gamma_{6n,1}-\Gamma_{6n,2}+\Gamma_{6n,3}-\Gamma_{6n,4}-\Gamma_{6n,5}+\Gamma_{6n,6}-\Gamma_{6n,7}+\Gamma_{6n,8}$, where $\Gamma_{6n,1}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}[(\hat{r}_i-r_i)\circ r_0]'[(\hat{r}_i-r_i)\circ(\hat{r}_0-r_0)]a_i$, $\Gamma_{6n,2}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}[(\hat{r}_i-r_i)\circ r_0]'\left[(\hat{r}_{1i}-r_{1i})\circ(\hat{r}_{2i}-r_{2i})\right]a_i$, $\Gamma_{6n,3}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left[r_i\circ(\hat{r}_0-r_0)\right]'\left[(\hat{r}_i-r_i)\circ(\hat{r}_0-r_0)\right]a_i$, $\Gamma_{6n,4}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left[r_i\circ(\hat{r}_0-r_0)\right]'[(\hat{r}_{1i}-r_{1i})\circ(\hat{r}_{2i}-r_{2i})]a_i$, $\Gamma_{6n,5}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left[(\hat{r}_{1i}-r_{1i})\circ r_{2i}\right]'\left[(\hat{r}_i-r_i)\circ(\hat{r}_0-r_0)\right]a_i$, $\Gamma_{6n,6}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}[(\hat{r}_{1i}-r_{1i})\circ r_{2i}]'\left[(\hat{r}_{1i}-r_{1i})\circ(\hat{r}_{2i}-r_{2i})\right]a_i$, $\Gamma_{6n,7}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left[r_{1i}\circ(\hat{r}_{2i}-r_{2i})\right]'\left[(\hat{r}_i-r_i)\circ(\hat{r}_0-r_0)\right]a_i$, and $\Gamma_{6n,8}\equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left[r_{1i}\circ(\hat{r}_{2i}-r_{2i})\right]'\left[(\hat{r}_{1i}-r_{1i})\circ(\hat{r}_{2i}-r_{2i})\right]a_i$. By Lemma B.3(b) and (B.2),

$$|\Gamma_{6n,1}|\leq\|r_0\|\,\|\hat{r}_0-r_0\|\,\bar{\Gamma}_{3n,1}=O_p\left(\nu_{bc}+n^{-1/2}b^{-1}\right)O_p(b^{\frac{d}{2}+2}(b^{-d-2}+c^{-d-1}))=o_P(1).$$

By Cauchy-Schwarz inequality, Lemma B.3(c), and equations (B.2) and (B.3),

$$\begin{aligned}
|\Gamma_{6n,2}| &\leq \|r_0\| \sup_{y\in\mathcal{Y}_0} \|\hat{r}_2(y) - r_2(y)\| b^{\frac{d}{2}+2} \sum_{i=1}^{n} \|\hat{r}_i - r_i\| \|\hat{r}_{1i} - r_{1i}\| a_i \\
&\leq \|r_0\| \sup_{y\in\mathcal{Y}_0} \|\hat{r}_2(y) - r_2(y)\| \left(\bar{\Gamma}_{3n,1}\bar{\Gamma}_{3n,2}\right)^{1/2} \\
&= O_P\left(\nu_{bc} + n^{-1/2}b^{-1}\sqrt{\ln n}\right) \left\{O_P\left(b^{\frac{d}{2}+2}\left(b^{-d-2} + c^{-d-1}\right)\right) O_P\left(b^{\frac{d}{2}+2}\left(b^{-d_x-2} + c^{-d_x}\right)\right)\right\}^{1/2} \\
&= o_P(1).
\end{aligned}$$

Note that $\Gamma_{6n,3} = b^{\frac{d}{2}+2}\left((\hat{r}_0 - r_0) \circ (\hat{r}_0 - r_0)\right)' \bar{\Gamma}_{2n,5}$ where $\bar{\Gamma}_{2n,5} \equiv \sum_{i=1}^{n}\left((\hat{r}_i - r_i) \circ r_i\right) a_i$. By Lemma B.3(b) and the proof of Lemma B.6(a),

$$\Gamma_{6n,3} = b^{\frac{d}{2}+2} O_P\left(\nu_{bc} + n^{-1/2}b^{-1}\right) O_P(b^{-d-1} + n^{1/2}b^{-1} + c^{-d-1}) = o_P(1).$$

Note that $\Gamma_{6n,4} = (\hat{r}_0 - r_0)' \bar{\Gamma}_{6n,4}$ where $\bar{\Gamma}_{6n,4} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left(r_i \circ (\hat{r}_{1i} - r_{1i}) \circ (\hat{r}_{2i} - r_{2i})\right) a_i$. Following the proof of Lemma B.6(f), we can show that $\bar{\Gamma}_{6n,4} = o_P(1)$. This, in conjunction with Lemma B.3(b), implies that $\Gamma_{6n,4} = o_P(1)$. Note that $\Gamma_{6n,5} = (\hat{r}_0 - r_0)'\bar{\Gamma}_{2n,6}$ where $\bar{\Gamma}_{2n,6} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left((\hat{r}_i - r_i) \circ (\hat{r}_{1i} - r_{1i}) \circ r_{2i}\right) a_i$. By Lemma B.3 and the proof of Lemma B.6(b), $\Gamma_{6n,5} = O_P\left(\nu_{bc} + n^{-1/2}b^{-1}\right)\left(O_P(b^{(d_z-d_x)/2}) + o_P(1)\right) = o_P(1)$. Note that $|\Gamma_{6n,6}| \leq \sup_{y\in\mathcal{Y}_0} \|\hat{r}_2(y) - r_2(y)\| \bar{\Gamma}_{6n,6}$ where $\bar{\Gamma}_{6n,6} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}((\hat{r}_{1i} - r_{1i}) \circ r_{2i} \circ (\hat{r}_{1i} - r_{1i}))a_i$. Analogously to the proof of Lemma B.4, we can show that $\bar{\Gamma}_{6n,6} = O_P(b^{\frac{d}{2}+2}(b^{-d_x-2} + c^{-d_x}))$. Combining this with Lemma B.3(c) yields

$$|\Gamma_{6n,6}| = O_P\left(\nu_{bc} + n^{-1/2}b^{-1}\sqrt{\ln n}\right) O_P\left(b^{\frac{d}{2}+2}\left(b^{-d_x-2} + c^{-d_x}\right)\right) = o_P(1).$$

Observe that $\Gamma_{6n,7} \equiv (\hat{r}_0 - r_0)' \bar{\Gamma}_{2n,7}$ where $\bar{\Gamma}_{2n,7} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left[(\hat{r}_i - r_i) \circ r_{1i} \circ (\hat{r}_{2i} - r_{2i})\right] a_i$. By Lemma B.3(b) and the proof of B.6(c), $\Gamma_{6n,7} = O_P(\nu_{bc} + n^{-1/2}b^{-1}\sqrt{\ln n}) o_P(1) = o_P(1)$. Lastly, by Cauchy-Schwarz inequality and the study of $\Gamma_{2n,4}$ and $\Gamma_{3n,2}$ above $|\Gamma_{6n,8}| \leq \{\Gamma_{2n,4}\Gamma_{3n,2}\}^{1/2} = o_P(1)$. Consequently we have proved $\Gamma_{6n} = o_P(1)$. ■

Remark. Admittedly, the formulae for the asymptotic bias and variance are quite complicated because we consider the general local polynomial regressions to estimate both $f(y|x,z)$ and $D_x F(y|x,z)$ and each of the four estimates $\hat{r}_i$, r_0, $\hat{r}_{1i}$, and $\hat{r}_{2i}$ contribute to the asymptotic bias and variance of our test statistic in different manners.

Lemma B.4 *Suppose Assumptions C.1-C.7 hold. Then* $\Gamma_{2n,1} - \mathbb{B}_{1n} \xrightarrow{d} N\left(0, \sigma_0^2\right)$ *where* $\mathbb{B}_{1n} = n^{-1}b^{\frac{d}{2}+2}\sum_{i=1}^{n} \varphi(W_i, W_i) = O_P(b^{\frac{d}{2}+2}\left(b^{-d-2} + c^{-d-1}\right))$.

Proof. Recall $\mathbf{1}_i \equiv \mathbf{1}\{Y_i \in \mathcal{Y}_0\}$. Let $f_i \equiv f(Y_i|V_i)$ and $D_{xi} \equiv D_x F(Y_i|X_i, Z_i)$. By Lemma B.3(a), $\hat{r}_i - r_i = b^{-1}e_1\bar{\mathbf{S}}_b(V_i)^{-1}\mathbf{V}_b(Y_i; V_i) f_i^{-1}\mathbf{1}_i - D_{xi}e_2'\bar{\mathbf{S}}_c(V_i)^{-1}\mathbf{V}_c^{(L)}(Y_i; V_i) f_i^{-2}\mathbf{1}_i + O_P(\nu_{bc})$. It follows that

$$\begin{aligned}
\Gamma_{2n,1} &= b^{\frac{d}{2}+2}\sum_{i=1}^{n} \|(\hat{r}_i - r_i) \circ r_0\|^2 a_i \\
&= b^{\frac{d}{2}+2}\sum_{i=1}^{n} \left\|\left(b^{-1}e_1\bar{\mathbf{S}}_b(V_i)^{-1}\mathbf{V}_b(Y_i; V_i) f_i^{-1} - D_{xi}e_2'\bar{\mathbf{S}}_c(V_i)^{-1}\mathbf{V}_c^{(L)}(Y_i; V_i) f_i^{-2}\right) \circ r_0\right\|^2 a_i \\
&\quad + nb^{\frac{d}{2}+2} O_P\left(\nu_{bc}^2 + \nu_{bc}\left(\nu_{1b}b^{-1} + b^p + \nu_{2c} + c^{p+1} + c^r\right)\right) \\
&= \bar{\Gamma}_{2n,1} + o_P(1),
\end{aligned}$$

where $\bar{\Gamma}_{2n,1} \equiv b^{\frac{d}{2}+2} \sum_{i=1}^{n} \left\| \left(b^{-1} e_1 \bar{\mathbf{S}}_b (V_i)^{-1} \mathbf{V}_b (Y_i; V_i) f_i^{-1} - D_{xi} e_2' \bar{\mathbf{S}}_c (V_i)^{-1} \mathbf{V}_c^{(L)} (Y_i; V_i) f_i^{-2} \right) \circ r_0 \right\|^2 a_i$,[11] and we use the fact that $\mathbf{1}_i a_i = a_i$ as $a(y; v)$ has compact support $\mathcal{Y}_0 \times \mathcal{V}_0$. Let $\zeta_k(w) \equiv \zeta_k(y; v)$ be as defined in Section 3.2. Then

$$\bar{\Gamma}_{2n,1} = b^{\frac{d}{2}+2} \sum_{i=1}^{n} \left\| n^{-1} \sum_{k=1}^{n} \zeta_k(W_i) \circ r_0 \right\|^2 a_i = n^{-2} b^{\frac{d}{2}+2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \zeta(W_{i_1}, W_{i_2}, W_{i_3}),$$

where $\zeta(W_{i_1}, W_{i_2}, W_{i_3}) \equiv \left(\zeta_{i_2}(W_{i_1}) \circ r_0\right)' \left(\zeta_{i_3}(W_{i_1}) \circ r_0\right) a_{i_1}$. Let $\varphi(w_{i_1}, w_{i_2}) \equiv E[\zeta(W_1, w_{i_1}, w_{i_2})]$, and $\bar{\zeta}(w_{i_1}, w_{i_2}, w_{i_3}) \equiv \zeta(w_{i_1}, w_{i_2}, w_{i_3}) - \varphi(w_{i_2}, w_{i_3})$. We can decompose $\bar{\Gamma}_{2n,1}$ as $\bar{\Gamma}_{2n,1} = \bar{\Gamma}_{2n,11} + \bar{\Gamma}_{2n,12}$, where

$$\bar{\Gamma}_{2n,11} = n^{-1} b^{\frac{d}{2}+2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \varphi(W_{i_1}, W_{i_2}) \text{ and } \bar{\Gamma}_{2n,12} = n^{-2} b^{\frac{d}{2}+2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \bar{\zeta}(W_{i_1}, W_{i_2}, W_{i_3}).$$

Consider $\bar{\Gamma}_{2n,12}$ first. Write $E(\bar{\Gamma}_{2n,12}^2) = n^{-4} b^{d+4} \sum_{i_1, \ldots, i_6}^{n} E\left[\bar{\zeta}(W_{i_1}, W_{i_2}, W_{i_3}) \bar{\zeta}(W_{i_4}, W_{i_5}, W_{i_6})\right]$. Noting that $E\left[\bar{\zeta}(W_{i_1}, w_{i_2}, w_{i_3})\right] = E\left[\bar{\zeta}(w_{i_1}, W_{i_2}, w_{i_3})\right] = E[\bar{\zeta}(w_{i_1}, w_{i_2}, W_{i_3})] = 0$, $E[\bar{\zeta}(W_{i_1}, W_{i_2}, W_{i_3}) \bar{\zeta}(W_{i_4}, W_{i_5}, W_{i_6})] = 0$ if there are more than three distinct elements in $\{i_1, \ldots, i_6\}$. With this, it is easy to show that $E(\bar{\Gamma}_{2n,12}^2) = O(n^{-1} b^{d+4} (b^{-4-3d} + c^{-3(d+1)})) = o(1)$. Hence $\bar{\Gamma}_{2n,12} = o_P(1)$ by the Chebyshev inequality.

For $\bar{\Gamma}_{2n,11}$, we have $\bar{\Gamma}_{2n,11} = n^{-1} b^{\frac{d}{2}+2} \sum_{i=1}^{n} \varphi(W_i, W_i) + 2n^{-1} b^{\frac{d}{2}+2} \sum_{1 \le i < j \le n} \varphi(W_i, W_j) \equiv \mathbb{B}_{1n} + \mathbb{V}_{1n}$, say, where $\varphi(W_i, W_j) = \int \zeta(w, W_i, W_j) \, dF(w) = \int (\zeta_i(w) \circ r_0)' \left(\zeta_j(w) \circ r_0\right) a(w) \, dF(w)$, and $\mathbb{B}_{1n}$ and $\mathbb{V}_{1n}$ contribute to the asymptotic bias and variance of $\bar{\Gamma}_{2n,11}$, respectively. Note that as $\mathbb{V}_{1n}$ is a second-order degenerate U-statistic, we can easily verify that all the conditions of Theorem 1 of Hall (1984) are satisfied and a central limit theorem applies to it: $\mathbb{V}_{1n} \xrightarrow{d} N\left(0, \sigma_0^2\right)$, where $\sigma_0^2 = \lim_{n \to \infty} \sigma_n^2$ and $\sigma_n^2 = 2b^{d+4} E\left[\varphi(W_1, W_2)\right]^2$. Thus $\Gamma_{2n,11} - \mathbb{B}_{1n} \xrightarrow{d} N\left(0, \sigma_0^2\right)$.

Lastly, noting that $E|\mathbb{B}_{1n}| = b^{\frac{d}{2}+2} O\left(b^{-d-2} + c^{-d-1}\right)$, we have $\mathbb{B}_{1n} = O_P(b^{\frac{d}{2}+2}(b^{-d-2} + c^{-d-1}))$ by Markov inequality. ■

Lemma B.5 *Suppose Assumptions C.1-C.7 hold. Then*

(a) $\Gamma_{2n,2} = b^{\frac{d}{2}+2} \sum_{i=1}^{n} \|r_i \circ (\hat{r}_0 - r_0)\|^2 a_i = o_P(1)$,

(b) $\Gamma_{2n,3} = b^{\frac{d}{2}+2} \sum_{i=1}^{n} \|(\hat{r}_{1i} - r_{1i}) \circ r_{2i}\|^2 a_i = \mathbb{B}_{2n} + o_P(1)$,

(c) $\Gamma_{2n,4} = b^{\frac{d}{2}+2} \sum_{i=1}^{n} \|r_{1i} \circ (\hat{r}_{2i} - r_{2i})\|^2 a_i = o_P(1)$,

where $\mathbb{B}_{2n} \equiv n^{-4} b^{\frac{d}{2}+2} \sum_{i=1}^{n} \left\| \sum_{j=1}^{n} \sum_{k=1}^{n} \zeta_k(Y_j; X_i, Z_j) \circ r_{2i} \right\|^2 a_i = O_P(b^{\frac{d}{2}+2}\left(b^{-d_x-2} + c^{-d_x}\right))$. *If* $d_z > 0$, *then (b) also holds when we replace* $\mathbb{B}_{2n}$ *by* $\bar{\mathbb{B}}_{2n} \equiv n^{-4} b^{\frac{d}{2}+2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\| \sum_{k=1}^{n} \zeta_j(Y_k; X_i, Z_k) \circ r_{2i} \right\|^2 a_i$.

Proof. (a) Note that $\Gamma_{2n,2} \le n b^{\frac{d}{2}+2} \|\hat{r}_0 - r_0\|^2 \bar{\Gamma}_{2n,2}$ where $\bar{\Gamma}_{2n,2} \equiv n^{-1} \sum_{i=1}^{n} \|r_i\|^2 a_i$. By Assumptions C.2(ii) and C.3(i), the compact support of a, and Markov inequality, $\bar{\Gamma}_{2n,2} = O_P(1)$. Using this and Lemma B.3(b) we have $\Gamma_{2n,2} = n b^{\frac{d}{2}+2} O_P\left(\nu_{bc}^2 + n^{-1} b^{-2}\right) O_P(1) = o_P(1)$.

(b) Noting that $\hat{r}_1(x) - r_1(x) = \frac{1}{n} \sum_{i=1}^{n} [\hat{r}(Y_i; x, Z_i) - r(Y_i; x, Z_i)] + \frac{1}{n} \sum_{i=1}^{n} [r(Y_i; x, Z_i) - r_1(x)]$, we have $b^{\frac{d}{2}+2} \sum_{i=1}^{n} \|[\hat{r}_1(X_i) - r_1(X_i)] \circ r_{2i}\|^2 a_i = R_{3n} + R_{4n} + 2R_{5n}$, where

$$R_{3n} \equiv \frac{b^{\frac{d}{2}+2}}{n^2} \sum_{i=1}^{n} \left\| \sum_{j=1}^{n} [\hat{r}(Y_j; X_i, Z_j) - r(Y_j; X_i, Z_j)] \circ r_{2i} \right\|^2 a_i,$$

[11] Write $\bar{\Gamma}_{2n,1} = b^{\frac{d}{2}+2} \sum_{i=1}^{n} \|(\xi_{1ni} - \xi_{2ni}) \circ r_0\|^2 a_i$, where $\xi_{1ni} \equiv b^{-1} e_1 \bar{\mathbf{S}}_b (V_i)^{-1} \mathbf{V}_b (Y_i; V_i) f_i^{-1}$ and $\xi_{2ni} \equiv D_{xi} e_2' \bar{\mathbf{S}}_c (V_i)^{-1} \mathbf{V}_c^{(L)} (Y_i; V_i) f_i^{-2}$. By straightforward moment calculations, we can show that ξ_{1ni} contributes to both the asymptotic bias and variance of the test statistic whereas ξ_{2ni} only contributes to the asymptotic bias.

$$R_{4n} \equiv \frac{b^{\frac{d}{2}+2}}{n^2}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}\left[r\left(Y_j;X_i,Z_j\right)-r\left(X_i\right)\right]\circ r_{2i}\right\|^2 a_i, \text{ and}$$

$$R_{5n} \equiv \frac{b^{\frac{d}{2}+2}}{n^2}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}\left[\hat{r}\left(Y_j;X_i,Z_j\right)-r\left(Y_j;X_i,Z_j\right)\right]\circ r_{2i}\right\|\left\|\sum_{k=1}^{n}\left[r\left(Y_k;X_i,Z_k\right)-r\left(X_i\right)\right]\circ r_{2i}\right\| a_i.$$

By Lemma B.3(a) we can readily show that $R_{3n}=\mathbb{B}_{2n}+o_P\left(1\right)$. We further decompose $\mathbb{B}_{2n}$ as $\mathbb{B}_{2n}=\mathbb{B}_{2n,1}+\mathbb{B}_{2n,2}$, where

$$\mathbb{B}_{2n,1} \equiv \frac{b^{\frac{d}{2}+2}}{n^4}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\left(\zeta_{i_3}\left(Y_{i_2};X_{i_1},Z_{i_2}\right)\circ r_{2i_1}\right)'\left(\zeta_{i_3}\left(Y_{i_4};X_{i_1},Z_{i_4}\right)\circ r_{2i_1}\right)a_{i_1} \text{ and}$$

$$\mathbb{B}_{2n,2} \equiv \frac{b^{\frac{d}{2}+2}}{n^4}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\sum_{i_5=1,i_5\neq i_3}^{n}\left(\zeta_{i_3}\left(Y_{i_2};X_{i_1},Z_{i_2}\right)\circ r_{2i_1}\right)'\left(\zeta_{i_5}\left(Y_{i_4};X_{i_1},Z_{i_4}\right)\circ r_{2i_1}\right)a_{i_1}.$$

By direct moment calculations and the Chebyshev inequality, we can show that $\mathbb{B}_{2n,2}=O_P(b^{\frac{d_z}{2}}+b^{\frac{d}{2}+2}c^{-\frac{d_x}{2}})$ which is $o_P\left(1\right)$ under Assumption A.7 if $d_z>0$, and that

$$\mathbb{B}_{2n,1}=\bar{\mathbb{B}}_{2n}=\frac{1}{n^4}b^{\frac{d}{2}+2}\sum_{i_1=1}^{n}\sum_{i_3=1}^{n}\left\|\sum_{i_2=1}^{n}\zeta_{i_3}\left(Y_{i_2};X_{i_1},Z_{i_2}\right)\circ r_{2i_1}\right\|^2 a_{i_1}=O_P(b^{\frac{d}{2}+2}(b^{-d_x-2}+c^{-d_x})).$$

It follows that $R_{3n}=\bar{\mathbb{B}}_{2n}+o_P\left(1\right)$ if $d_z>0$. By Markov inequality, $R_{4n}=O_P(b^{\frac{d}{2}+2})=o_P\left(1\right)$. By Cauchy-Schwarz inequality, $R_{5n}\leq\left\{R_{3n}R_{4n}\right\}^{1/2}=O_P(\{[b^{\frac{d}{2}+2}(b^{-d_x-2}+c^{-d_x})+1]b^{\frac{d}{2}+2}\}^{1/2})=o_P\left(1\right)$. This completes the proof of part (b).

(c) Noting that $\hat{r}_2\left(y\right)-r_2\left(y\right)=\frac{1}{n}\sum_{i=1}^{n}\left[\hat{r}\left(y;X_i,Z_i\right)-r\left(y;X_i,Z_i\right)\right]+\frac{1}{n}\sum_{i=1}^{n}\left[r\left(y;X_i,Z_i\right)-r_2\left(y\right)\right]$, by Cauchy-Schwarz inequality we have $\Gamma_{2n,4}\leq 2R_{6n}+2R_{7n}$, where $R_{6n}\equiv\frac{b^{\frac{d}{2}+2}}{n^2}\sum_{i=1}^{n}||\sum_{j=1}^{n}r_{1i}\circ[\hat{r}\left(Y_i;X_j,Z_j\right)$ $-r\left(Y_i;X_j,Z_j\right)]||^2a_i$ and $R_{7n}\equiv\frac{b^{\frac{d}{2}+2}}{n^2}\sum_{i=1}^{n}||\sum_{j=1}^{n}r_{1i}\circ\left[r\left(Y_i;X_j,Z_j\right)-r_2\left(Y_i\right)\right]||^2a_i$. By Markov inequality $R_{7n}=O_P(b^{\frac{d}{2}+2})=o_P\left(1\right)$. For R_{6n} we can first apply Lemma B.3 to show that $R_{6n}=\bar{R}_{6n}+o_P\left(1\right)$, where

$$\bar{R}_{6n}\equiv\frac{b^{\frac{d}{2}+2}}{n^2}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}r_{1i}\circ\left[b^{-1}e_1\bar{\mathbf{S}}_b\left(V_j\right)^{-1}\mathbf{V}_b\left(Y_i;V_j\right)f_{ij}^{-1}-f_{ij}^{-2}D_{xij}e_2'\bar{\mathbf{S}}_c\left(X_j,Z_j\right)^{-1}\mathbf{V}_c^{(L)}\left(Y_i;V_j\right)\right]\right\|^2 a_i,$$

$f_{ij}\equiv f\left(Y_i|V_j\right)$, and $D_{xij}\equiv D_xF\left(Y_i|V_j\right)$. Observe that $\bar{R}_{6n}\leq 2\bar{R}_{6n,1}+2\bar{R}_{6n,2}$, where

$$\bar{R}_{6n,1} = \frac{b^{\frac{d}{2}+2}}{n^2}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}r_{1i}\circ\left(b^{-1}e_1\bar{\mathbf{S}}_b\left(V_j\right)^{-1}\mathbf{V}_b\left(Y_i;V_j\right)f_{ij}^{-1}\right)\right\|^2 a_i, \text{ and}$$

$$\bar{R}_{6n,2} = \frac{b^{\frac{d}{2}+2}}{n^2}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}r_{1i}\circ\left(f_{ij}^{-2}D_{xij}e_2'\bar{\mathbf{S}}_c\left(V_j\right)^{-1}\mathbf{V}_c^{(L)}\left(Y_i;V_j\right)\right)\right\|^2 a_i.$$

By straightforward but tedious moment calculations, we can show that

$$\begin{aligned}E\left(\bar{R}_{6n,1}\right) &= \frac{b^{\frac{d}{2}}}{n^4}\sum_{i=1}^{n}E\left\|\sum_{j=1}^{n}\sum_{k=1}^{n}r_{1i}\circ\left(e_1\bar{\mathbf{S}}_b\left(V_j\right)^{-1}\mu_b\left(V_k-V_j\right)K_b\left(V_k-V_j\right)\bar{\mathbf{1}}_{Y_i}\left(W_k\right)f_{ij}^{-1}\right)\right\|^2 a_i\\ &= O\left(b^{\frac{d}{2}}+n^{-1}b^{-\frac{d}{2}}+n^{-2}b^{-\frac{3d}{2}}\right)=o\left(1\right).\end{aligned}$$

Similarly, $E\left(\bar{R}_{6n,2}\right) = b^{\frac{d}{2}+2}O\left(1+n^{-1}c^{-(d+1)}+n^{-2}c^{-2(d+1)}\right) = o(1)$. Then $\bar{R}_{6n} = o_P(1)$ by Markov inequality. It follows that $R_{6n} = o_P(1)$ and $\Gamma_{2n,4} = o_P(1)$. ■

Lemma B.6 *Suppose that Assumptions C.1-C.7 hold. Then*

(a) $\Gamma_{2n,5} = 2b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left((\hat{r}_i - r_i)\circ r_0\right)'\left(r_i \circ (\hat{r}_0 - r_0)\right)a_i = o_P(1)$,

(b) $\Gamma_{2n,6} = -2b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left((\hat{r}_i - r_i)\circ r_0\right)'\left((\hat{r}_{1i} - r_{1i})\circ r_{2i}\right)a_i = -2\mathbb{B}_{3n} + o_P(1)$,

(c) $\Gamma_{2n,7} = -2b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left((\hat{r}_i - r_i)\circ r_0\right)'\left(r_{1i} \circ (\hat{r}_{2i} - r_{2i})\right)a_i = o_P(1)$,

(d) $\Gamma_{2n,8} = -2b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left(r_i \circ (\hat{r}_0 - r_0)\right)'\left((\hat{r}_{1i} - r_{1i})\circ r_{2i}\right)a_i = o_P(1)$,

(e) $\Gamma_{2n,9} = -2b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left(r_i \circ (\hat{r}_0 - r_0)\right)'\left(r_{1i} \circ (\hat{r}_{2i} - r_{2i})\right)a_i = o_P(1)$,

(f) $\Gamma_{2n,10} = 2b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left((\hat{r}_{1i} - r_{1i})\circ r_{2i}\right)'\left(r_{1i} \circ (\hat{r}_{2i} - r_{2i})\right)a_i = o_P(1)$,

where $\mathbb{B}_{3n} \equiv n^{-3}b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left[\left(\sum_{l=1}^{n}\zeta_l(W_i)\circ r_0\right)'\left(\sum_{j=1}^{n}\sum_{k=1}^{n}\zeta_k(Y_j;X_i,Z_j)\circ r_2(Y_i)\right)\right]a_i = O_P(b^{(d_z-d_x)/2} + b^{\frac{d}{2}+2}c^{-d_x})$.

Proof. (a) Noting that $\Gamma_{2n,5} = 2b^{\frac{d}{2}+2}\left((\hat{r}_0 - r_0)\circ r_0\right)'\bar{\Gamma}_{2n,5}$ where $\bar{\Gamma}_{2n,5} = \sum_{i=1}^{n}\left((\hat{r}_i - r_i)\circ r_i\right)a_i$. By Lemma B.3(a), we can show that $\bar{\Gamma}_{2n,5} = \bar{\Gamma}_{2n,51} + o_P(n^{1/2}b^{-(\frac{d}{2}+1)})$, where

$$\begin{aligned}
\bar{\Gamma}_{2n,51} &= \sum_{i=1}^{n}\left[\left(b^{-1}e_1\bar{\mathbf{S}}_b(V_i)^{-1}\mathbf{V}_b(Y_i;V_i)f_i^{-1} - D_{xi}e_2'\bar{\mathbf{S}}_c(V_i)^{-1}\mathbf{V}_c^{(L)}(Y_i;V_i)f_i^{-2}\right)\circ r_i\right]a_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{n}\zeta_{1j}(Y_i;V_i)\circ r_i\right]a_i - \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{n}\zeta_{2j}(Y_i;V_i)\circ r_i\right]a_i \equiv R_{8n} + R_{9n}.
\end{aligned}$$

In view of $R_{8n} = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{n}[b^{-1}e_1\bar{\mathbf{S}}_b(V_i)^{-1}\mu_b(V_j - V_i)K_b(V_j - V_i)\bar{\mathbf{1}}_{Y_i}(W_j)f_i^{-1}]\circ r_i\right]a_i$, it is easy to show that $E\|R_{8n}\|^2 = O\left(nb^{-2} + b^{-2d-2}\right)$. Thus $R_{8n} = O_P\left(b^{-d-1} + n^{1/2}b^{-1}\right)$. Similarly, $R_{9n} = O_P(c^{-d-1} + n^{1/2})$. It follows that $\bar{\Gamma}_{2n,51} = O_P(b^{-d-1} + n^{1/2}b^{-1} + c^{-d-1})$. Then by Lemma B.3(b) and the fact that $\nu_{bc} = O_P\left(n^{-1/2}b^{-1}\right)$, $\Gamma_{2n,5} = b^{\frac{d}{2}+2}O_P\left(n^{-1/2}b^{-1}\right)\left[O_P\left(b^{-d-1} + n^{1/2}b^{-1} + c^{-d-1}\right) + o_P(n^{1/2}b^{-(\frac{d}{2}+1)})\right] = o_P(1)$.

(b) Write $\Gamma_{2n,6} = -2r_0'\bar{\Gamma}_{2n,6}$ where $\bar{\Gamma}_{2n,6} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left((\hat{r}_i - r_i)\circ(\hat{r}_{1i} - r_{1i})\circ r_{2i}\right)a_i$. Then $\bar{\Gamma}_{2n,6} = R_{10n} + R_{11n}$, where $R_{10n} \equiv n^{-1}b^{\frac{d}{2}+2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left((\hat{r}_i - r_i)\circ[\hat{r}(Y_j;X_i,Z_j) - r(Y_j;X_i,Z_j)]\circ r_{2i}\right)a_i$ and $R_{11n} \equiv n^{-1}b^{\frac{d}{2}+2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left((\hat{r}_i - r_i)\circ[r(Y_j;X_i,Z_j) - r_1(X_i)]\circ r_{2i}\right)a_i$. Using Lemma B.3, we can show that $R_{10n} = \bar{R}_{10n} + o_P(1)$, where

$$\begin{aligned}
\bar{R}_{10n} &= n^{-1}b^{\frac{d}{2}+2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big\{\left[b^{-1}e_1\bar{\mathbf{S}}_b(V_i)^{-1}\mathbf{V}_b(Y_i;V_i)f_i^{-1} - D_{xi}e_2'\bar{\mathbf{S}}_c(V_i)^{-1}\mathbf{V}_c^{(L)}(Y_i;V_i)f_i^{-2}\right] \\
&\quad \circ\left[b^{-1}e_1\bar{\mathbf{S}}_b(X_i,Z_j)^{-1}\mathbf{V}_b(Y_j;X_i,Z_j)f_{ji}^{-1} - D_{xi}e_2'\bar{\mathbf{S}}_c(X_i,Z_j)^{-1}\mathbf{V}_c^{(L)}(Y_j;X_i,Z_j)f_{ji}^{-2}\right]\circ r_{2i}\Big\}a_i\mathbf{1}_i\mathbf{1}_j \\
&= n^{-3}b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left[\sum_{l=1}^{n}\zeta_l(W_i)\circ\sum_{j=1}^{n}\sum_{k=1}^{n}\zeta_k(Y_j;X_i,Z_j)\circ r_2(Y_i)\right]a_i.
\end{aligned}$$

Noting that $E\left\|\bar{R}_{10n}\right\|^2 = O\left(b^{d+4}\left(b^{-2d_x-4} + c^{-2d_x}\right)\right)$, we have $\bar{R}_{10n} = O_P(b^{(d_z-d_x)/2} + b^{\frac{d}{2}+2}c^{-d_x})$ which is $o_P(1)$ under Assumption A.7 if $d_z > d_x$ and otherwise not. Hence $r_0'R_{10n} = \mathbb{B}_{3n} + o_P(1)$ and $\mathbb{B}_{3n} = O_P(b^{(d_z-d_x)/2} + b^{\frac{d}{2}+2}c^{-d_x})$. For R_{11n}, we apply Cauchy-Schwarz inequality to obtain $R_{11n} \le \{\alpha_{1n}\alpha_{2n}\}^{1/2}$, where $\alpha_{1n} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^{n}\|(\hat{r}_i - r_i)\circ r_{2i}\|^2 a_i$ and $\alpha_{2n} \equiv n^{-1}b^{\frac{d}{2}+2}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}[r(Y_j;X_i,Z_j) - r_1(X_i)]\right\|^2 a_i$. Analogously to the determination of the probability order of $\Gamma_{2n,1}$, we can show that $\alpha_{1n} = O_P(b^{\frac{d}{2}+2}(b^{-d-2}+$

$c^{-d-1})$). Next, $\alpha_{2n} = O_P(b^{\frac{d}{2}+2})$ by Markov inequality. It follows that $R_{11n} = O_P(b^{-\frac{d}{4}} + b^{\frac{d}{4}+1}c^{-\frac{d+1}{2}})O_P(b^{\frac{d}{4}+1})$ $= O_P(b + b^{\frac{d}{2}+2}c^{-\frac{d+1}{2}}) = o_P(1)$ and $\Gamma_{2n,6} = -2\mathbb{B}_{3n} + o_P(1)$.

(c) Write $\Gamma_{2n,7} = -2r_0'\bar{\Gamma}_{2n,7}$ where $\bar{\Gamma}_{2n,7} \equiv b^{\frac{d}{2}+2}\sum_{i=1}^n[(\hat{r}_i - r_i)\circ r_{1i}\circ(\hat{r}_{2i} - r_{2i})]a_i$. We further decompose $\bar{\Gamma}_{2n,7}$ as $\bar{\Gamma}_{2n,7} = R_{12n} + R_{13n}$, where $R_{12n} \equiv n^{-1}b^{\frac{d}{2}+2}\sum_{i=1}^n\sum_{j=1}^n\{(\hat{r}_i - r_i)\circ r_{1i}\circ[\hat{r}(Y_i;V_j) - r(Y_i;V_j)]\}a_i$ and $R_{13n} \equiv n^{-1}b^{\frac{d}{2}+2}\sum_{i=1}^n\sum_{j=1}^n\{(\hat{r}_i - r_i)\circ r_{1i}\circ[r(Y_i;V_j) - r_2(Y_i)]\}a_i$. Following the analysis of R_{10} and R_{11n}, we can readily show that $R_{sn} = o_P(1)$ for $s = 12, 13$. It follows that $\Gamma_{2n,7} = o_P(1)$.

(d) Noting that $\Gamma_{2n,8} = 2b^{\frac{d}{2}+2}(\hat{r}_0 - r_0)'\bar{\Gamma}_{2n,8}$ where $\bar{\Gamma}_{2n,8} \equiv \sum_{i=1}^n(r_i\circ(\hat{r}_{1i} - r_{1i})\circ r_{2i})a_i$. Then $\bar{\Gamma}_{2n,8} = R_{14n} + R_{15n}$, where $R_{14n} \equiv \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\{r_i\circ[\hat{r}(Y_j;X_i,Z_j) - r(Y_j;X_i,Z_j)]\circ r_{2i}\}a_i$ and $R_{15n} \equiv \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\{r_i\circ[r(Y_j;X_i,Z_j) - r_1(X_i)]\circ r_{2i}\}a_i$. By straightforward moment calculations, we can show that $R_{15n} = O_P(n^{1/2})$. By Lemma B.3(a), we can show that $R_{14n} = \bar{R}_{14n} + o_P(n^{1/2}b^{-(\frac{d}{2}+1)})$, where

$$\bar{R}_{14n} = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n[r_i\circ\zeta_k(Y_j;X_i,Z_j)\circ r_{2i}]a_i.$$

Noting that $E\|\bar{R}_{14n}\|^2 = O(nb^{-2} + (b^{-d_x-2} + b^{-2d_z-2} + c^{-d_x-1} + c^{-2d_z-2}) + n^{-1}(b^{-d-2-d_z} + c^{-d-2-d_z}))$, we have $\bar{R}_{14n} = O_P(n^{1/2}b^{-1} + b^{-d_z-1} + c^{-d_z-1})$. Consequently, $\bar{\Gamma}_{2n,8} = O_P(n^{1/2}b^{-1} + b^{-d_z-1} + c^{-d_z-1}) + o_P(n^{1/2}b^{-(\frac{d}{2}+1)})$. Then by Lemma B.3(b) and the fact that $\nu_{bc} = O_P(n^{-1/2}b^{-1})$,

$$\begin{aligned}\Gamma_{2n,8} &= b^{\frac{d}{2}+2}O_P\left(n^{-1/2}b^{-1}\right)\left[O_P\left(n^{1/2}b^{-1} + b^{-d_z-1} + c^{-d_z-1}\right) + o_P(n^{1/2}b^{-(\frac{d}{2}+1)})\right]\\ &= O_P(b^{\frac{d}{2}} + n^{-1/2}b^{\frac{d_x-d_z}{2}} + n^{-1/2}b^{\frac{d}{2}+2}c^{-d_z-1}) + o_P(1) = o_P(1).\end{aligned}$$

(e) By the Cauchy-Schwarz inequality and Lemmas B.5(a) and (c), $|\Gamma_{2n,9}| \le 2(\Gamma_{2n,2}\Gamma_{2n,4})^{1/2} = o_P(1)$.

(f) Let $\bar{\Delta}r_{1ij} = \hat{r}(Y_j;X_i,Z_j) - r(Y_j;X_i,Z_j)$, $\Delta r_{1ij} = r(Y_j;X_i,Z_j) - r_1(X_i)$, $\bar{\Delta}r_{2ij} = \hat{r}(Y_i;X_j,Z_j) - r(Y_i;X_j,Z_j)$, and $\Delta r_{2ij} = r(Y_i;X_j,Z_j) - r_2(Y_i)$. Using $\hat{r}_{1i} - r_{1i} = \frac{1}{n}\sum_{j=1}^n(\bar{\Delta}r_{1ij} + \Delta r_{1ij})$ and $\hat{r}_{2i} - r_{2i} = \frac{1}{n}\sum_{k=1}^n(\bar{\Delta}r_{2ik} + \Delta r_{2ik})$, we can decompose $\Gamma_{2n,10}$ as follows:

$$\begin{aligned}\Gamma_{2n,10} &= 2n^{-2}b^{\frac{d}{2}+2}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\Big\{(\Delta r_{1ij}\circ r_{2i})'(r_{1i}\circ\Delta r_{2ik})a_i + (\bar{\Delta}r_{1ij}\circ r_{2i})'(r_{1i}\circ\bar{\Delta}r_{2ik})a_i\\ &\qquad + (\bar{\Delta}r_{1ij}\circ r_{2i})'(r_{1i}\circ\Delta r_{2ik})a_i + (\Delta r_{1ij}\circ r_{2i})'(r_{1i}\circ\bar{\Delta}r_{2ik})a_i\Big\}\\ &\equiv 2R_{16n} + 2R_{17n} + 2R_{18n} + 2R_{19n},\ \text{say}.\end{aligned}$$

By moment calculations, $E(R_{16n}) = O(b^{\frac{d}{2}+2})$ and $E(R_{16n}^2) = O(b^{d+4})$, implying that $R_{16n} = O_P(b^{\frac{d}{2}+2}) = o_P(1)$. For R_{17n}, we can show that $R_{17n} = \bar{R}_{17n} + o_P(1)$, where

$$\bar{R}_{17n} = n^{-4}b^{\frac{d}{2}+2}\sum_{i_1=1}^n\sum_{i_2=1}^n\sum_{i_3=1}^n\sum_{i_4=1}^n\sum_{i_5=1}^n(\zeta_{i_4}(Y_{i_2};X_{i_1},Z_{i_2})\circ r_{2i_1})'(r_{1i_1}\circ\zeta_{i_5}(Y_{i_1};X_{i_3},Z_{i_3}))a_{i_1}.$$

Noting that $E(\bar{R}_{17n}^2) = O(b^{d+4}b^{-4}) + n^{-1}b^{d+4}(b^{-d_x-4} + b^{-d_z-4} + c^{-d_x} + c^{-d_z})) = o(1)$, we have $\bar{R}_{17n} = o_P(1)$. Similarly, we can show that $R_{18n} = o_P(1)$ and $R_{19n} = o_P(1)$. Consequently, $\Gamma_{2n,10} = o_P(1)$. ■

Proof of Theorem 3.2. The proof follows closely from that of Theorems 3.1. By (B.1) and the proof of Theorem 3.1. Now $\hat{\Gamma} = n^{-1}b^{-(\frac{d}{2}+2)}\Gamma_{1n} + n^{-1}b^{-(\frac{d}{2}+2)}\Gamma_{4n} + n^{-1}b^{-(\frac{d}{2}+2)}\Gamma_{5n} + o_P(1)$. It is easy to show that $n^{-1}b^{-(\frac{d}{2}+2)}\Gamma_{1n} = n^{-1}\sum_{i=1}^n\|r_i\circ r_0 - r_{1i}\circ r_{2i}\|^2a_i = \mu_A + o_P(1)$ and $n^{-1}b^{-(\frac{d}{2}+2)}\Gamma_{sn} = o_P(1)$ under $\mathbb{H}_A$ for $s = 4, 5$. In addition, under $\mathbb{H}_A$, we have $n^{-1}b^{-(\frac{d}{2}+2)}\hat{\mathbb{B}}_n = o_P(1)$ and $\hat{\sigma}_n^2 \xrightarrow{p} \sigma_A^2$. It follows that $n^{-1}b^{-(\frac{d}{2}+2)}T_n = n^{-1}b^{-(\frac{d}{2}+2)}[nb^{(\frac{d}{2}+2)}\hat{\Gamma} - \hat{\mathbb{B}}_n]/\sqrt{\hat{\sigma}_n^2} = \mu_A/\sigma_A + o_P(1)$ and the result follows. ■

Proof of Theorem 3.3. By assumption, $\delta_n(x,\cdot)$ is strictly increasing and continuously differentiable for each $x \in \mathcal{X}_n$. Let $\delta_n^{-1}(x,\cdot)$ denote the inverse function of $\delta_n(x,\cdot)$. Let $\delta_{n,u}(x,\cdot)$ denote the derivative of $\delta_n(x,u)$ with respect to $\cdot$, i.e., $\delta_{n,u}(x,u) = 1 + \gamma_n \delta_u(x,u)$ where $\delta_u(x,u) = \partial\delta_u(x,u)/\partial u$. Then by the inverse function theorem

$$\frac{d\delta_n^{-1}(x,t)}{dt} = \frac{1}{\delta_{n,u}\left(x,\delta_n^{-1}(x,t)\right)} = \frac{1}{1+\gamma_n\delta_u\left(x,\delta_n^{-1}(x,t)\right)}.$$

It follows that

$$\delta_n^{-1}(x,t) = \eta_n(x) + \int_0^t \frac{1}{1+\gamma_n\delta_u\left(x,\delta_n^{-1}(x,s)\right)}ds$$

for some function $\eta_n(x)$ that does not depend on t. Noting that $\frac{1}{1+a} = 1 - a + o(a)$ when $a = o(1)$, we have

$$\begin{aligned}\delta_n^{-1}(x,t) &= \eta_n(x) + \int_0^t \left[1-\gamma_n\delta_u\left(x,\delta_n^{-1}(x,s)\right)\right]ds + o(\gamma_n t) \\ &= \eta_n(x) + t - \gamma_n\bar{\eta}_n(x,t) + o(\gamma_n t),\end{aligned} \tag{B.4}$$

where $\bar{\eta}_n(x,t) = \int_0^t \delta_u\left(x,\delta_n^{-1}(x,s)\right)ds$.

Let $F_{U_n|Z_n}(\cdot,z)$ and $f_{U_n|Z_n}(\cdot,z)$ denote the conditional CDF and PDF of U_{ni} given $Z_{ni}=z$, respectively. Let $t(x,y) \equiv G^{-1}(y) - H_1(x)$. Then by (3.2), (B.4) and the strict monotonicity of G and δ_n,

$$\begin{aligned}F_n(y \mid x,z) &\equiv \Pr\left[Y_{ni} \le y \mid X_{ni}=x, Z_{ni}=z\right] \\ &= \Pr\left[G\left[H_1(X_{ni}) + U_{ni} + \gamma_n\delta(X_{ni},U_{ni})\right] \le y \mid X_{ni}=x, Z_{ni}=z\right] \\ &= \Pr\left[\delta_n(x,U_{ni}) \le t(x,y) \mid Z_{ni}=z\right] \\ &= \Pr\left[U_{ni} \le \delta_n^{-1}(x,t(x,y)) \mid Z_{ni}=z\right] \\ &= F_{U_n|Z_n}\left[\delta_n^{-1}(x,t(x,y)),z\right] \\ &= F_{U_n|Z_n}\left[\eta_n(x)+t(x,y)-\gamma_n\bar{\eta}_n(x,t(x,y))+o(\gamma_n t(x,y)),z\right] \\ &\simeq F_{U_n|Z_n}\left[\eta_n(x)+t(x,y),z\right] + f_{U_n|Z_n}\left[\eta_n(x)+t(x,y),z\right]\left[-\gamma_n\bar{\eta}_n(x,t(x,y))+o(\gamma_n t(x,y))\right] \\ &\simeq F_{U_n|Z_n}\left[\eta_n(x)+t(x,y),z\right] - \gamma_n\varsigma_n(y;x,z)\end{aligned}$$

where $\simeq$ is used to indicate that we have suppressed high-order remainder terms, and $\varsigma_n(y;x,z) \equiv f_{U_n|Z_n}\left[\eta_n(x)+t(x,y),z\right]\bar{\eta}_n(x,t(x,y))$. Let $F_{1,U_n|Z_n}$ be the derivative of $F_{U_n|Z_n}$ with respect to its first argument. Then

$$\begin{aligned}r_n^0(y;x,z) &\equiv \frac{\partial F_n(y \mid x,z)/\partial x}{\partial F_n(y \mid x,z)/\partial y} \\ &\simeq \frac{F_{1,U_n|Z_n}\left[G^{-1}(y)-H_1(x),z\right]\cdot\{\partial\left[\eta_n(x)-H_1(x)\right]/\partial x\} - \gamma_n\varsigma_{n,x}(y;x,z)}{F_{1,U_n|Z_n}\left[G^{-1}(y)-H_1(x),z\right]\cdot\left[\partial G^{-1}(y)/\partial y\right] - \gamma_n\varsigma_{n,y}(y;x,z)} \\ &\simeq \frac{\partial\left[\eta_n(x)-H_1(x)\right]/\partial x}{\partial G^{-1}(y)/\partial y} + \gamma_n\Delta_n(y;x,z),\end{aligned} \tag{B.5}$$

where $\varsigma_{n,x}(y;x,z) = \partial\varsigma_n(y;x,z)/\partial x$, $\varsigma_{n,y}(y;x,z) = \partial\varsigma_n(y;x,z)/\partial y$,

$$\Delta_n(y;x,z) = \frac{\varsigma_{n,y}(y;x,z)\{\partial\left[\eta_n(x)-H_1(x)\right]/\partial x\}}{F_{1,U_n|Z_n}\left[G^{-1}(y)-H_1(x),z\right]\cdot\left[\partial G^{-1}(y)/\partial y\right]^2} - \frac{\varsigma_{n,x}(y;x,z)}{F_{1,U_n|Z_n}\left[G^{-1}(y)-H_1(x),z\right]\cdot\left[\partial G^{-1}(y)/\partial y\right]},$$

and the second $\simeq$ in (B.5) follows from the fact that $\frac{A-a}{B-b} \simeq \frac{A}{B} - \frac{a}{B} + \frac{Ab}{B^2}$ when $a = o(1)$, $B = o(1)$ and B is bounded away from 0. So the functions s_{1n} and s_2 exist and are given by

$$s_{1n}(x) = \frac{C\partial[\eta_n(x) - H_1(x)]}{\partial x} \quad \text{and} \quad s_2(y) = \frac{1}{C\partial G^{-1}(y)/\partial y},$$

where $C \neq 0$ is an arbitrary constant. Clearly, $s_2 : \mathbb{R} \to \mathbb{R}_+$ if $C > 0$ and $s_2 : \mathbb{R} \to \mathbb{R}_-$ if $C < 0$. The measurable functions S_{1n} and S_2 are given by $C[\eta_n(x) - H_1(x)]$ and CG^{-1}, respectively. ■

Proof of Corollary 3.4. Let $\mathbf{1}_{ni} = \mathbf{1}\{Y_{ni} \in \mathcal{Y}_0\}$ and $\mathbf{1}_y = \mathbf{1}\{y \in \mathcal{Y}_0\}$. By Theorem 3.3,

$$\begin{aligned} r_{0n} &= E[s_{1n}(X_{ni})] E[s_2(Y_{ni})\mathbf{1}_{ni}] + \gamma_n E_{Y_n} E_{X_n Z_n}[\Delta_n(Y_{ni}; X_{ni}, Z_{ni})\mathbf{1}_{ni}] + o(\gamma_n), \\ r_{1n}(x) &= s_{1n}(x) E[s_2(Y_{ni})\mathbf{1}_{ni}] + \gamma_n E[\Delta_n(Y_{ni}; x, Z_{ni})\mathbf{1}_{ni}] + o(\gamma_n), \\ r_{2n}(y) &= E[s_{1n}(X_{ni})] s_2(y)\mathbf{1}_y + \gamma_n E[\Delta_n(y; X_{ni}, Z_{ni})]\mathbf{1}_y + o(\gamma_n) \end{aligned}$$

It follows that $r_n(y; x, z) \circ r_{0n} - r_{1n}(x) \circ r_{2n}(y) = \gamma_n \bar{\Delta}_n(y; x, z) + o(\gamma_n)$ for all $(y, x, z) \in \mathcal{W}_{0n}$, where

$$\begin{aligned} \bar{\Delta}_n(y; x, z) &= \{\Delta_n(y; x, z)\mathbf{1}_y\} \circ \{E[s_{1n}(X_{ni})] E[s_2(Y_{ni})\mathbf{1}_{ni}]\} \\ &\quad + \{s_{1n}(x) s_2(y)\mathbf{1}_y\} \circ \{E_{Y_n} E_{X_n Z_n}[\Delta_n(Y_{ni}; X_{ni}, Z_{ni})\mathbf{1}_{ni}]\} \\ &\quad - \{s_{1n}(x) E[s_2(Y_{ni})\mathbf{1}_{ni}]\} \circ \{E[\Delta_n(y; X_{ni}, Z_{ni})]\mathbf{1}_y\} \\ &\quad - \{E[\Delta_n(Y_{ni}; x, Z_{ni})\mathbf{1}_{ni}] \circ E[s_{1n}(X_{ni})] s_2(y)\}\{\mathbf{1}_y\}. \blacksquare \end{aligned} \tag{B.6}$$

Proof of Theorem 3.5. The proof follows closely from that of Theorem 3.1, now keeping the additional terms that do not vanish under $\mathbb{H}_A(\gamma_n)$ with $\gamma_n = n^{-1/2} b^{-\frac{d}{4}-1}$. Let $r_{ni} \equiv r_n(Y_{ni}; X_{ni}, Z_{ni})$, $r_{1ni} \equiv r_{1n}(X_{ni})$, $r_{2ni} \equiv r_{2n}(Y_{ni})$, $\hat{r}_i \equiv \hat{r}(Y_{ni}; X_{ni}, Z_{ni})$, $\hat{r}_{1i} \equiv \hat{r}_1(X_{ni})$, and $\hat{r}_{2i} \equiv \hat{r}_2(Y_{ni})$. Noting that $\hat{\mathbb{B}}_n = \mathbb{B}_n + o_P(1)$ and $\hat{\sigma}_n^2 = \sigma_0^2 + o_P(1)$ under $\mathbb{H}_A(\gamma_n)$, it suffices to show that under $\mathbb{H}_A(\gamma_n)$, (i) $\Gamma_{1n} \xrightarrow{p} \mu_0$, (ii) $\Gamma_{4n} = o_p(1)$ and (iii) $\Gamma_{5n} = o_p(1)$, where Γ_{1n}, Γ_{4n}, and Γ_{5n} are defined after (B.1) with r_0, r_i, r_{1i}, and r_{2i} now replaced by r_{n0}, r_{ni}, r_{1ni}, and r_{2ni}, respectively. The results in previous lemmas continue to hold when the single-index IID observations (Y_i, X_i, Z_i) are replaced by the double-array IID observations (Y_{ni}, X_{ni}, Z_{ni}). Let $V_{ni} \equiv (X_{ni}', Z_{ni}')'$ and $f_{ni} \equiv f_n(Y_{ni}|V_{ni})$.

(i) holds under $\mathbb{H}_A(\gamma_n)$ because $\Gamma_{1n} = b^{\frac{d}{2}+2} \sum_{i=1}^n \|r_{ni} \circ r_{n0} - r_{1ni} \circ r_{2ni}\|^2 a_i = n^{-1} \sum_{i=1}^n ||\bar{\Delta}_n(Y_{ni}; X_{ni}, Z_{ni})||^2 a_i = \mu_0 + o_P(1)$ by the law of large numbers. For (ii), we decompose Γ_{4n} as $\Gamma_{4n} = \Gamma_{4n,1} + \Gamma_{4n,2} - \Gamma_{4n,3} - \Gamma_{4n,4}$, where

$$\begin{aligned} \Gamma_{4n,1} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n (r_{ni} \circ r_{0n} - r_{1ni} \circ r_{2ni})' ((\hat{r}_i - r_{ni}) \circ r_{0n}) a_i, \\ \Gamma_{4n,2} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n (r_{ni} \circ r_{0n} - r_{1ni} \circ r_{2ni})' (r_{ni} \circ (\hat{r}_0 - r_{0n})) a_i, \\ \Gamma_{4n,3} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n (r_{ni} \circ r_{0n} - r_{1ni} \circ r_{2ni})' ((\hat{r}_{1i} - r_{1ni}) \circ r_{2ni}) a_i, \\ \Gamma_{4n,4} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n (r_{ni} \circ r_{0n} - r_{1ni} \circ r_{2ni})' (r_{1ni} \circ (\hat{r}_{2i} - r_{2ni})) a_i. \end{aligned}$$

It suffices to prove $\Gamma_{4n,s} = o_P(1)$ for $s = 1, 2, 3, 4$. We only prove $\Gamma_{4n,1} = o_P(1)$ as the other cases are similar. Let $\bar{\Delta}_{ni} \equiv \bar{\Delta}_n(Y_{ni}; X_{ni}, Z_{ni})$. Under $\mathbb{H}_A(\gamma_n)$ we apply Lemma B.3(a) to obtain

$$\Gamma_{4n,1} = n^{-\frac{1}{2}} b^{\frac{d}{4}+1} \sum_{i=1}^n \delta_{ni}' ((\hat{r}_i - r_{ni}) \circ r_{0n}) a_i = \bar{\Gamma}_{4n,1} + n^{\frac{1}{2}} b^{\frac{d}{4}+1} O_P(\nu_{bc}) = \bar{\Gamma}_{4n,1} + o_P(1),$$

where $\bar{\Gamma}_{4n,1} \equiv n^{-\frac{1}{2}} b^{\frac{d}{4}+1} \sum_{i=1}^{n} \bar{\Delta}_{ni}' \{ [b^{-1} e_1 \bar{\mathbf{S}}_b (V_{ni})^{-1} \mathbf{V}_b (Y_{ni}; X_{ni}) f_{ni}^{-1} - D_{xi} e_2' \bar{\mathbf{S}}_c (V_{ni})^{-1} \mathbf{V}_c^{(L)} (V_{ni}) f_{ni}^{-2}] \circ r_{0n} \} a_i$. Write $\bar{\Gamma}_{4n,1} = n^{-\frac{3}{2}} b^{\frac{d}{4}+1} \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{\Delta}_{ni}' \left(\zeta_j (Y_{ni}; V_{ni}) \circ r_{0n} \right) a_i$. Then $E \left\| \bar{\Gamma}_{4n,1} \right\|^2 = O((b^{\frac{d}{2}} + n^{-1} b^{\frac{d}{2}+2} (b^{-d-2} + c^{-d-1}) + n^{-2} b^{\frac{d}{2}+2} (b^{-2d-2} + c^{-2d-1})) = o(1)$, implying that $\bar{\Gamma}_{4n,1} = o_P(1)$. It follows that $\Gamma_{4n,1} = o_P(1)$.

We now show (iii). Decompose $\Gamma_{5n} = \Gamma_{5n,1} - \Gamma_{5n,2}$ where $\Gamma_{5n,1} = (\hat{r}_0 - r_0)' \bar{\Gamma}_{5n,1}$, $\Gamma_{5n,2} = \gamma_n b^{\frac{d}{2}+2} \sum_{i=1}^{n} \bar{\Delta}_{ni}' [(\hat{r}_{1i} - r_{1ni}) \circ (\hat{r}_{2i} - r_{2ni})] a_i$, and $\bar{\Gamma}_{5n,1} \equiv b^{\frac{d}{2}+2} \sum_{i=1}^{n} [r_{ni} \circ r_{0n} - r_{1ni} \circ r_{2ni}] \circ (\hat{r}_i - r_{ni}) a_i$. Analogous to the study of $\Gamma_{4n,1}$, one can readily show that $\bar{\Gamma}_{5n,1} = o_P(1)$. Then by Lemma B.3(b), $\Gamma_{5n,1} = O_P(\nu_{bc} + n^{-1/2} b^{-1}) o_P(1) = o_P(1)$. Analogous to the proof of Lemma B.6(f), we can show that $\Gamma_{5n,2} = o_P(\gamma_n)$. Thus $\Gamma_{5n} = o_P(1)$.

Consequently, $P(T_n \geq z | \mathbb{H}_A(n^{-1/2} b^{-\frac{d}{4}-1})) \to 1 - \Phi(z - \mu_0/\sigma_0)$. This concludes the proof of the theorem. ■

C Proof of the main results in Section 5

Proof of Proposition 5.1. We first prove the "if" part. By the definition of the hazard function, for any values (y, x, ς) on the support of (Y, X, ξ), $h(y, x, \varsigma) = \frac{f(y|x,\varsigma)}{1 - F(y|x,\varsigma)}$, where $f(y|x, \varsigma)$ and $F(y|x, \varsigma)$ are conditional PDF and CDF of Y given $(X, \xi) = (x, \varsigma)$, respectively. Then,

$$
\begin{aligned}
F(y|x, \varsigma) &= P\left[G\left(H_1(X) + \ln\left(\frac{-\ln(1-\varepsilon)}{\xi} \right) \right) \leq y \,\middle|\, X = x, \xi = \varsigma \right] \\
&= P\left[\frac{-\ln(1-\varepsilon)}{\xi} \leq \exp\left[G^{-1}(y) - H_1(X) \right] \,\middle|\, X = x, \xi = \varsigma \right] \\
&= P\left[\varepsilon \leq 1 - \exp\left\{ -\varsigma \exp\left[G^{-1}(y) - H_1(x) \right] \right\} \right] \\
&= 1 - \exp\left\{ -\varsigma \exp\left[G^{-1}(y) - H_1(x) \right] \right\}.
\end{aligned}
$$

Thus $f(y|x, \varsigma) = \varsigma \exp\left\{ -\varsigma \exp\left[G^{-1}(y) - H_1(x) \right] \right\} \exp\left[G^{-1}(y) - H_1(x) \right] \frac{d\left[G^{-1}(y) \right]}{dy}$ and

$$
h(y, x, \varsigma) = \frac{f(y|x, \varsigma)}{1 - F(y|x, \varsigma)} = \left\{ \frac{\exp\left[G^{-1}(y) \right]}{g\left(G^{-1}(y) \right)} \right\} \exp\left[-H_1(x) \right] \varsigma = \lambda(y) \cdot \theta(x) \cdot \varsigma,
$$

where $\lambda(y) = \exp\left[G^{-1}(y) \right] / g\left(G^{-1}(y) \right)$, $g(s) = dG(s)/ds$, and $\theta(x) = \exp\left[-H_1(x) \right]$. This holds for all (y, x, ς) on the support of (Y, X, ξ), thus the "if" part is proved.

Next, we prove the "only if" part. Define the integrated hazard function $H(Y, X, \xi) = \int_0^Y h(y, X, \xi)\, dy$. Then

$$
H(Y, X, \xi) = \int_0^Y h(y, X, \xi)\, dy = \int_0^Y \lambda(y)\, dy \cdot \theta(X) \cdot \xi = \Lambda(Y) \cdot \theta(X) \cdot \xi,
$$

where $\Lambda(Y) = \int_0^Y \lambda(y)\, dy$. Let $F(Y \mid X, \xi)$ be the conditional CDF of Y given X and ξ. For any distribution function F, the integrated hazard function is related to its distribution function by $H(Y, X, \xi) = -\ln(1 - F(Y \mid X, \xi))$. Therefore $\Lambda(Y) \theta(X) \xi = -\ln(1 - F(Y \mid X, \xi))$. Define the random variable $\varepsilon = F(Y \mid X, \xi)$. By construction ε is uniformly distributed on $[0, 1]$ and $\varepsilon \perp (X, \xi)$ and $\Lambda(Y) \theta(X) \xi = -\ln(1 - \varepsilon)$. Thus $\ln[\Lambda(Y)] = \ln\left[\frac{-\ln(1-\varepsilon)}{\xi} \right] + \ln\left[\frac{1}{\theta(X)} \right]$. That is, $Y = G[H_1(X) + U]$ where $G(\cdot)$ is the inverse function of $\ln[\Lambda(\cdot)]$, $H_1(X) = -\ln[\theta(X)]$ and $U = \ln\left[\frac{-\ln(1-\varepsilon)}{\xi} \right]$. ■

REFERENCES

Abbring, J. H., Chiappori, P., Zavadil, T., 2008. Better safe than sorry? Ex ante and ex post moral hazard in dynamic insurance data. Working Paper, VU University Amsterdam and Columbia University.

Altonji, J., Matzkin, R., 2005. Cross section and panel data estimators for nonseparable models with endogenous regressors. Econometrica 73, 1053-1102.

Blundell, R., Powell, J., 2003. Endogeneity in nonparametric and semiparametric regression models. In Advances in Economics and Econometrics, eds. M. Dewatripont, L. Hansen, and S. Turnovsky, 294-311, Cambridge: Cambridge University Press.

Blundell, R., Powell, J., 2004. Endogeneity in semiparametric binary response models. Review of Economic Studies 71, 581-913.

Cameron, C., Trivedi, P., 2005. Microeconometrics: methods and applications. Cambridge University Press, New York.

Chiappori, P-A., Komunjer, I., Kristensen, D., 2013. Nonparametric identification and estimation of transformation models. Working paper, Department of Economics, Columbia University.

Cox, D. R., 1972. Regression models and life tables (with discussion). Journal of the Royal Statistical Society, Series B 34, 187-220.

Dawid, A. P., 1979. Conditional independence in statistical theory. Journal of the Royal Statistical Society, Series B 41, 1–31.

Ekeland, I., Heckman, J. J., Nesheim, L., 2004. Identification and estimation of hedonic models. Journal of Political Economy 112, 60-109.

Engle, R. F., 2000. The econometrics of ultra-high-frequency data. Econometrica 68, 1-22.

Fan, J., Jiang, J., 2005. Nonparametric inference for additive models. Journal of American Statistical Association 100, 890-907.

Fan, J., Yao, Q., Tong, H., 1996. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika 83, 189-206.

Fan, J., Zhang, C., Zhang, J., 2001. Generalized likelihood ratio statistics and Wilks phenomenon. Annals of Statistics 29, 153-193.

Fernandes, M., Grammig, J., 2005. Nonparametric specification tests for conditional duration models. Journal of Econometrics 127, 35-68.

Gozalo, P., Linton, O., 2001. Testing additivity in generalized nonparametric regression models with estimated parameters. Journal of Econometrics 104, 1-48.

Greene, W., 2011. Econometric Analysis. 7th edition. NJ: Prentice Hall.

Hall, P., 1984. Central limit theorem for integrated square error properties of multivariate nonparametric density estimators. Journal of Multivariate Analysis 14, 1-16.

Hansen, B. E., 2008. Uniform convergence rates for kernel estimation with dependent data. Econometric Theory 24, 726-748.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926-1947.

Heckman, J. J., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in econometric Models for duration data. Econometrica, 52, 271-320.

Heckman, J. J., Matzkin, R., Nesheim, L., 2005. Estimation and simulation of hedonic models. In T. Kehoe, T. Srinivasan, and J. Whalley (eds), Frontiers in Applied General Equilibrium, pp. 277-340. Cambridge University Press, Cambridge.

Heckman, J.J., and Robb, R., 1986. Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In: Wainer H., editor. Drawing Inferences from Self-Selected Samples. New York: Springer-Verlag; Mahwah, NJ: Lawrence Erlbaum Associates; pp. 63–107.

Hoderlein, S., Mammen, E., 2007. Identification of marginal effects in nonseparable models without monotonicity. Econometrica 75, 1513-1518.

Hoderlein, S., Su, L., White, H., 2011. Specification testing for nonparametric structural models with monotonicity in unobservables. Working paper, Dept. of Economics, Boston College.

Horowitz, J. L., 1996. Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. Econometrica 64, 103–137.

Horowitz, J. L., 2001. Nonparametric estimation of a generalized additive model with an unknown link function. Econometrica 69, 499–513.

Horowitz, J. L., 2014. Nonparametric additive models. In J. Racine, L. Su and A. Ullah (eds), The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics, pp. 129-148. Oxford University Press, Oxford.

Horowitz, J. L., Mammen, E., 2004. Nonparametric estimation of an additive model with a link function. Annals of Statistics 36, 2412–2443.

Horowitz, J. L., Mammen, E., 2007. Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. Annals of Statistics 35, 2589-2619.

Horowitz, J. L., Mammen, E., 2011. Oracle-efficient estimation of an additive model with an unknown link function. Econometric Theory 27, 582-608.

Ichimura, H., Lee, S., 2011. Identification and estimation of a nonparametric transformation model. Working paper, Department of Economics, Tokyo University.

Imbens, G., Newey, W., 2009. Identification and estimation of triangular simultaneous equations models without additivity. Econometrica 77, 1481-1512.

Jacho-Chávez, D., Lewbel, A., Linton, O., 2010. Identification and nonparametric estimation of a transformed additively separable model. Journal of Econometrics 156, 392-407.

Kennan, J. E., 1985. The duration of contract strikes in US manufacturing. Journal of Econometrics 28, 5-28.

Keifer, N. M., 1988. Economic duration data and hazard functions. Journal of Economic Literature 26, 646-679.

Kong, E., Linton, O., Xia, Y., 2010. Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. Econometric Theory 26, 1529-1564.

Lancaster, T., 1979. Econometric models for the duration of unemployment. Econometrica 47, 939-956.

Li, Q., Lu, X., Ullah, A., 2003. Multivariate local polynomial regression for estimating average derivatives. Journal of Nonparametric Statistics 15, 607-627.

Li, Q., Racine, J., 2003. Nonparametric estimation of distributions with categorical and continuous data. Journal of Multivariate Analysis, 86, 266-292.

Lu, X., White, H., 2013. Testing for separability in structural equations. Journal of Econometrics, forthcoming.

Mata, J., Portugal, P., 1994. Life duration of new firms. Journal of Industrial Economics, 42, 227-245.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency rates. Journal of Time Series Analysis 17, 571-599.

Matzkin, R. L., 2003. Nonparametric estimation of nonadditive random functions. Econometrica 71, 1339-1375.

Matzkin, R. L., 2007. Nonparametric identification. In J.J. Heckman and E.E. Leamer (eds), Handbook of Econometrics, Vol. 6, pp.5307-5368. Elsevier, B.V.

Politis, D., Romano, J., Wolf, M., 1999. Subsampling. New York: Springer-Verlag.

Ridder, G., 1990. The nonparametric identification of generalized accelerated failure-time models. Review of Economic Studies 57, 167-181.

Su, L., Tu, Y., Ullah, A., 2013. Testing additive separability of error term in nonparametric structural models. Econometric Reviews, forthcoming.

Van den Berg, G., 2001. Duration Models: Specification, Identification and Multiple Durations. In J.J. Heckman and E.E. Leamer (eds), Handbook of Econometrics, Vol. 5, pp.3381-3460. Elsevier, B.V.

White, H., Lu, X., 2011. Causal diagrams for treatment effect estimation with application to selection of efficient covariates. Review of Economics and Statistics 93, 1453-1459.